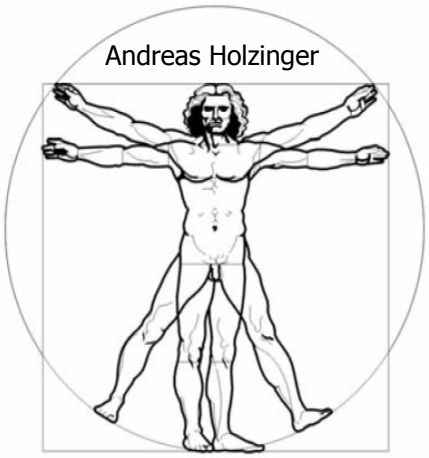
Andreas Holzinger



LV 706.046 3SE AK Mensch-Maschine Kommunikation

Usability Engineering for Software Developers

Applying User Centered Design

Schönen Nachmittag!

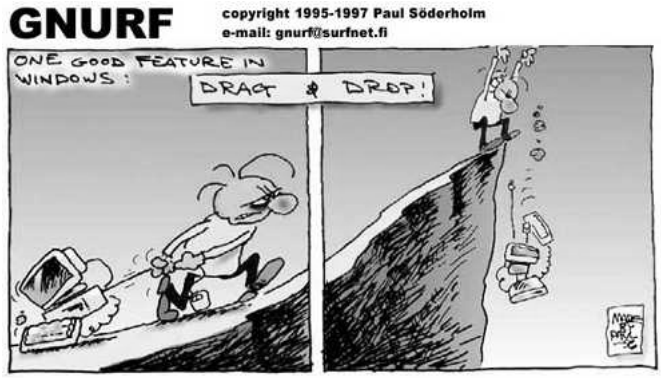
Revised schedule

- Work on the project together (today)
- Write a paper together (3 authors, finish today)
- Submit the 6 papers to Committee Chair (Extended Deadline: 6.6.2005) two copies (name and without)
  - via e-mail: andreas.holzinger@meduni-graz.at
- Peer review due to 15.6.2005
- e-Mail Check – all e-Mails correct?
- Written Final exam (30 %) 13.6.
- Presenting the paper in plenum 20.6.

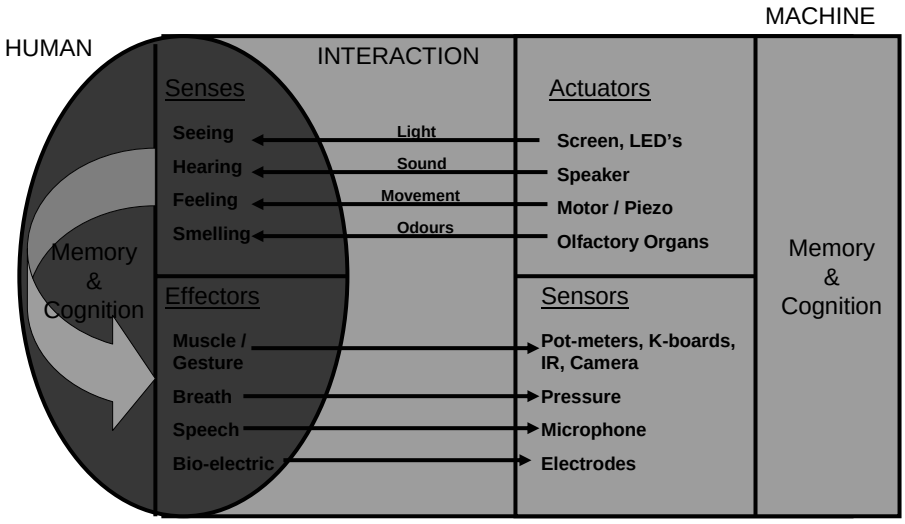
- Mini-Conference in i3
- 14:00 Keynote Martina Manhartsberger
- 6 technical sessions, every session will be chaired by a session-chair
- Usability Stammtisch with Keith Andrews in GRABA, Grazbachgasse (Beer !!!)

Remember: Students Material on
www.uni-graz.at/~holzinge/holzinger/usability.html

- How to write a good research paper and
- How to give a good research talk
- Checklist for Reviewers
- Common errors in English
- All lecture slides
- Videos



- Until today: rather WIMP orientated, but what about the Future ?



Bongers (2002) – Interaction Model based on Sensory Modalities

+ Multimodal Interfaces

+ progress of multimodal HCI

- Although progress has been impressive in multimodal human-computer interaction, we still face problems which we must solve before *ubiquitous* adoption can be realized

+ Example: Speech

- Speech Recognition = spoken words into computer text (e.g. medical report)
- Voice Recognition = spoken words produce execution of commands (e.g. in operating theatre)
- Both are a challenge for the field of Human-Computer Interaction & Usability Engineering ...

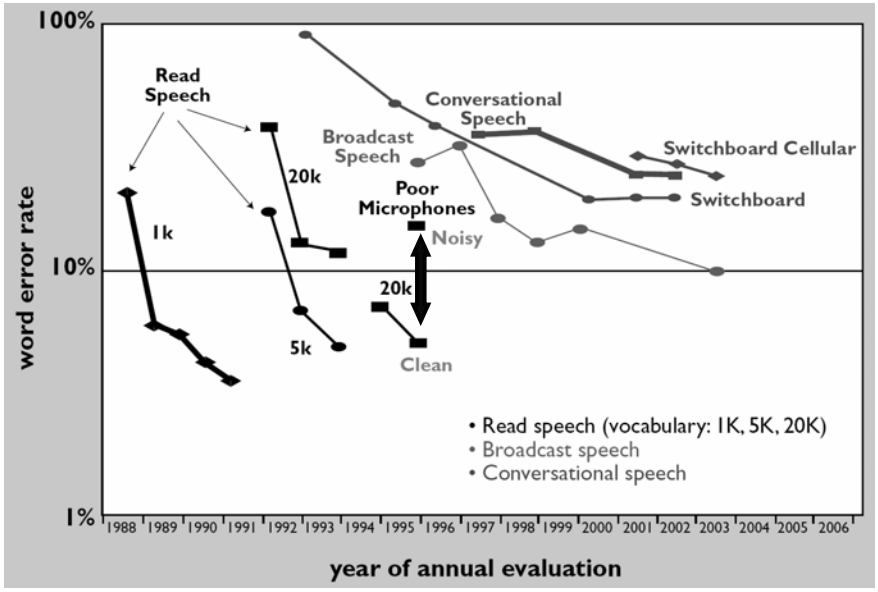
+ Kurzweil on speech software

Ray Kurzweil introduced the first commercial large-vocabulary speech recognition software in 1987

http://www.kurzweiltech.com/kai.html

Kurzweil, R. (1999), *The Age of Spiritual Machines:* Penguin, London.

+ Word error rate Deng & Huang (2004)

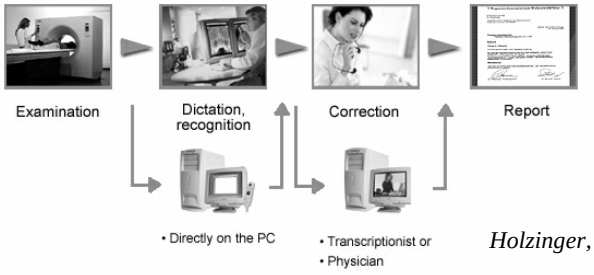


andreas.holzinger@meduni-graz.at

Deng & Huang (2004)

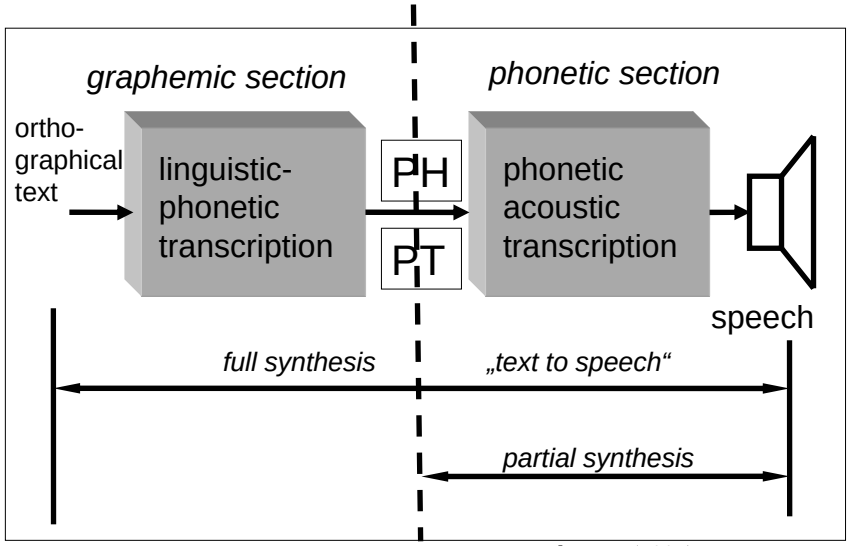
+ Usual setting: Text-to-Speech

	Mean Value (s)
Speech Recognition	139.02
Dictation to Tape	30.32
Hand written report	107.97

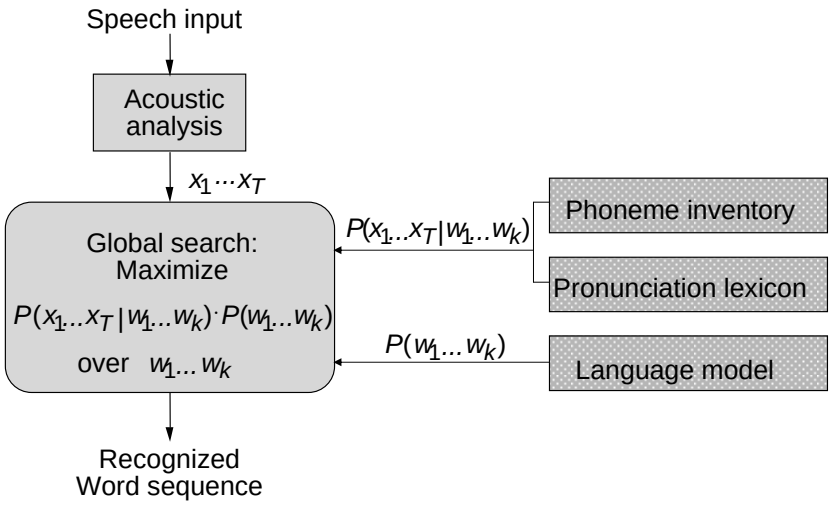


Holzinger, Ackerl, Searle, Sorantin (2004)

andreas.holzinger@meduni-graz.at

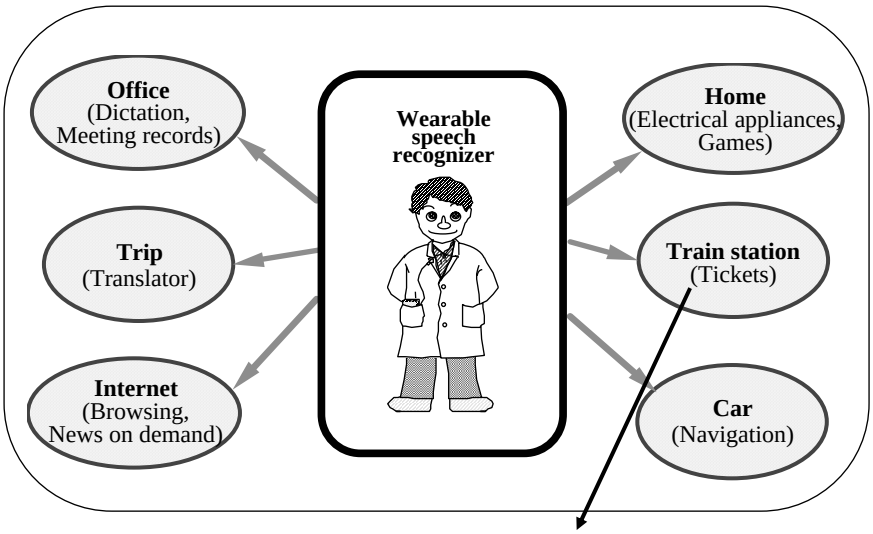


Holzinger (2002)



Holzinger (2002)

Ubiquitous computing environment



Beispiel Deutsche Bahn OSSI BAHNTICKET.wav

百聞不如一見

- bai wen bu ru yi jian
- Seeing it once is better than being told 100 times (Zhou Chongguo, Han Dynasty)
- or, A picture is worth ten thousand words Barnard (1927), often mis-quoted as an old Chinese proverb.

What I hear I forget, what I see I remember, what I do I understand.

这句话出自荀子《儒效篇》。

"不闻不若闻之，闻之不若见之，见之不若知之，知之不若行之；
学至于行之而止矣。"

http://www.chinapage.com/quote/quote.html

- Elliot Soloway: "A picture is worth a thousand words?" Ein Bild sagt mehr als tausend Worte

Plaue, C., Miller, T. & Stasko, J. (2004), Is a picture worth a thousand words? an evaluation of information awareness displays. Proceedings of: 2004 conference on Graphics interface, London, Ontario, Canada, 117-126.

Gregory, R.L., Eye and Brain, The Psychology of Seeing, Princeton University Press, 1990.

Gershon, N. D. (1996), Breaking the myth: one picture is NOT (always) worth a thousand words (panel).
Conference on Computer graphics and interactive techniques, 491-492.

- Is this valid for peephole displays?

Holzinger (2005)

+ text-to-speech

- End-users are very used to listen (especially mobile phone users);
- Audio information is a very natural way of information;
- It is location and eye-contact independent;
- Parallel tasks are possible;
- Simple ways for variations (transportation of meta information).

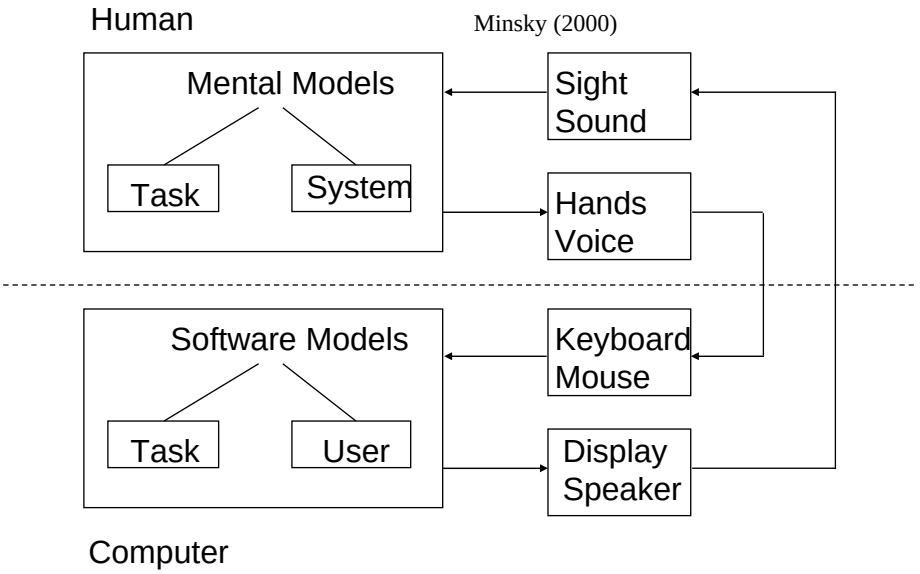
Witten (1982); Brewster (1998); Holzinger, Nischelwitzer & Sorantin (2005)

+ Speech UI Design

- Speech UIs allow
  - more natural computer access
  - computer use in more situations (e.g., hands free)
- Speech UIs are hard to get to work well because of
  - lack of visible state
  - tax working memory
  - recognition problems
  - natural language understanding is also a hard problem
- Multimodal UIs are
  - combination of two or more natural input modalities
    - e.g., speech & pen, speech & gesture, etc.
- Multimodal UIs address some of the problems by
  - helping disambiguate ambiguous inputs
  - helping corrections

Gentner & Stevens (1983)

Minsky (2000)



- **The user has control**
  - command languages
  - query languages
  - line editors
  - adventure games
  - ...
- **The computer has control**
  - Q&A
  - advisory systems
  - ...
- **Both have control (“conversation”)**
  - ?

```
$pwd
/usr/giovanna
$ls
filea fileb filec
$rm filea
$
```

(Unix)


```
GIVE THE NAMES OF ALL EMPLOYEES WHO HAVE JOBS WORKING AS
A SECRETARY IN THE CITY OF CHICAGO.
PRINT THE NAME OF ANY EMPLOYEE WITH CITY=CHICAGO AND
JOB=SECRETARY
THE NUMBER OF RECORDS TO RETRIEVE IS 30
SMITH
JONES
BROWN
...
```

(Intellect)

```
You are standing at the end of a road before a small
brick building. Around you is a forest. A small stream
flows down a gully.
ENTER BUILDING
You are inside a building, a well house for a large
spring. There are some keys on the ground here. There is
a shiny brass lamp nearby. There is a bottle of water
here.
GET WATER
Okay
GET KEYS
Okay
```

(Adventure, 1975)


```
(1) Patient's name: (first-last)
**FRED SMITH
(2) Sex:
**MALE
(3) Age:
**55
(4) Have you been able to obtain positive cultures
   from a site at which Fred Smith has an infection?
**YES
(5) What is the infection?
**PRIMARY-BACTEREMIA
(6) Please give the date and approximate time when
   signs of symptoms first appeared
```

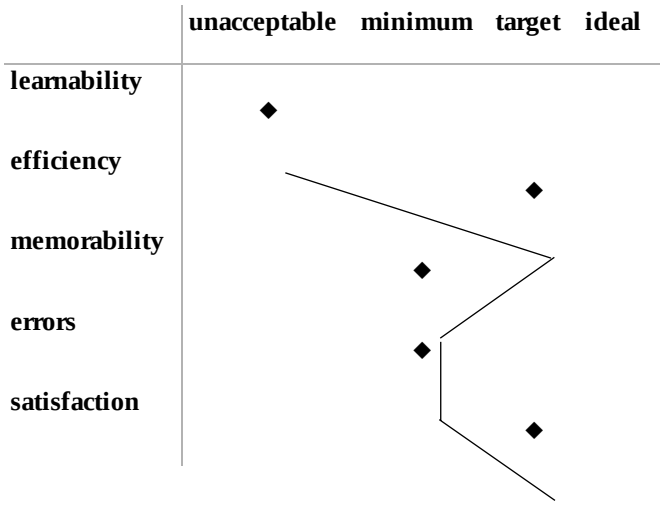
Shortliffe et al (1973): An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Comput. Biomed. Res. 6,* 544--560 (1973).

+ usability metrics

- Learnability
  - time to reach specified level of proficiency
    - e.g. complete a specified, representative task
  - note that learning is a continuum
- Memorability
  - test users on commands after trial session
- Errors
  - number of errors in completing specified task
- Subjective satisfaction
  - rating scales AND
  - physiological measures

  -> biological usability testing
- Efficiency
  - times for experts to complete specified task(s)
  - frequency of 'non-productive' actions
  - ratio of used to unused commands

+ Example: a usability profile



+ Jakob Nielsen's heuristics

1.0 – circa 1990	2.0 – circa 1994
Simple and natural dialog	Aesthetic and minimalist design
Speak the user's language	Match between system and real world
Minimize user memory load	Recognition rather than recall
Be consistent	Consistency and standards
Provide feedback	Visibility of system status
Provide clearly marked exits	User control and freedom
Provide shortcuts	Flexibility and efficiency of use
Provide good error messages	Help users recognize, diagnose, and recover from errors
Prevent errors	Error prevention
Help and documentation	Help and documentation

Nielsen, J. & Molich, R. (1990), Heuristic evaluation of user interfaces. Proceedings of: CHI 90, Seattle (WA), 249-256.

Nielsen, J. & Mack, R. L. (Eds.) (1994), *Usability Inspection Methods*, New York, Wiley.

+ Jakob Nielsen

- PHD in user interface design from the Technical University of Denmark
- Until 1998 he was a Sun Microsystems Distinguished Engineer.
- Principal of the Nielsen Norman Group which he co-founded with Don Norman (former VP of research at Apple).

http://www.useit.com/jakob

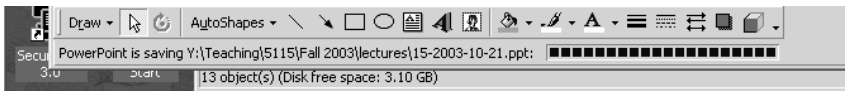
+ Jakob Nielsen

- Founded the "discount usability engineering" movement for fast and cheap improvements of user interfaces
- invented several usability methods, including Heuristic Evaluation
- 74 United States patents

- Papers:
- Nielsen, J. (1999). User interface directions for the Web. Communications of the ACM 42, 1 (January), 65-72.
- Nielsen, J., and Faber, J. M. (1996). Improving System Usability Through Parallel Design. IEEE Computer, 29, 2 (February), 29-35.
- Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. Proc. ACM CHI'94 Conf. (Boston, MA, April 24-28), 152-158.
- Nielsen, J., and Levy, J. (1994). Measuring usability - preference vs. performance. Communications of the ACM 37, 4 (April), 66-75.

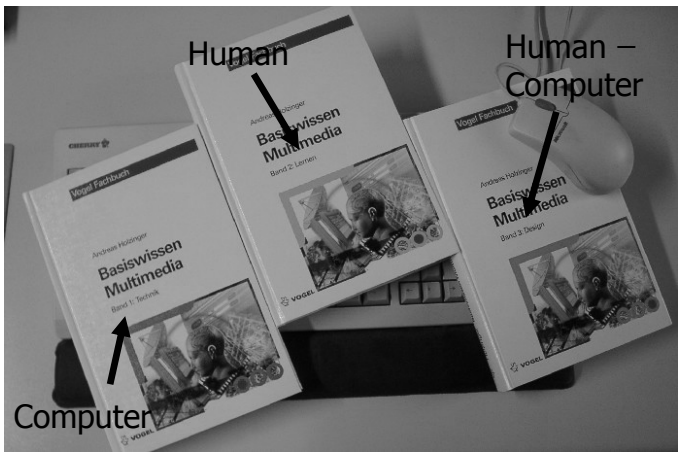
+ Example Heuristics

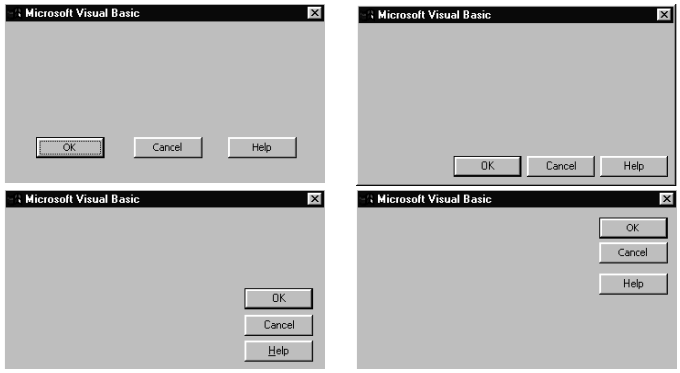
- H2-1: Visibility of system status
  - keep users informed about what is going on
  - example: pay attention to response time



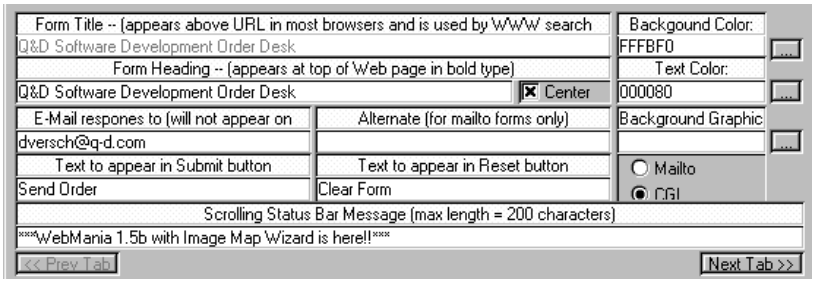
Response time	*Expectations*
1/10 s	Limit for immediate Response of the System (e.g. Keyboard - Screen)
1 s	Limit for proper Response of the System (e.g. click on a button)
2 s	Limit for feedback of a (still) running program (e.g. indiaction beam)
10 s	Limit for attention (Termination of the user)

www.basiswissen-multimedia.at





- H2-4: Consistency & standards

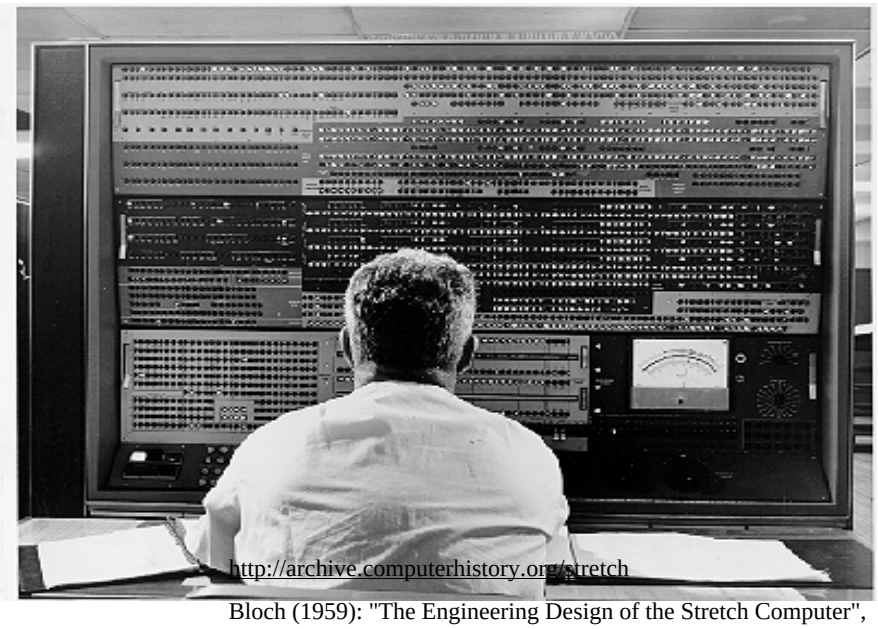


- H2-8: Aesthetic and minimalist design
  - no irrelevant information in dialogues

+ Participatory Heuristic

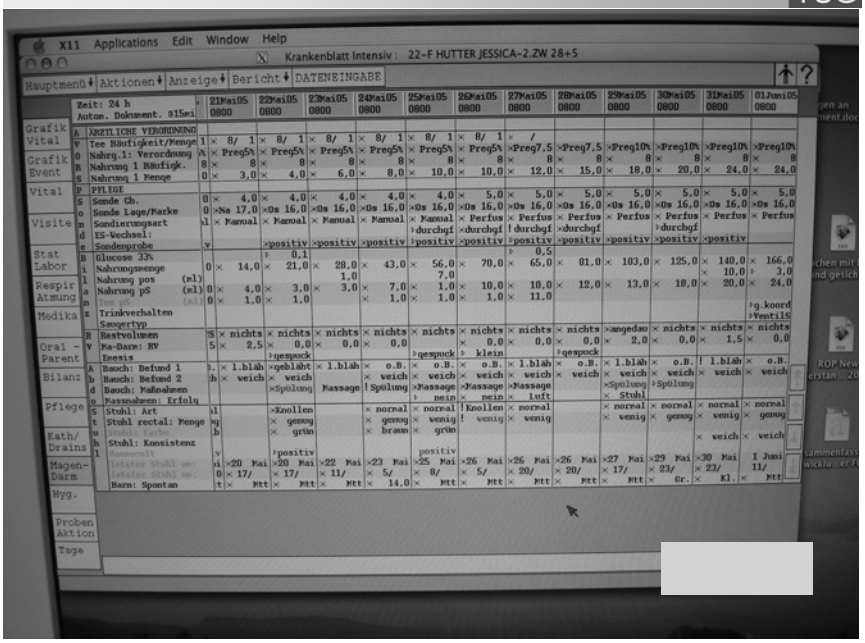
- Baker, K., Greenberg, S., Gutwin, C. (2002) From methods to design: Empirical development of a heuristic evaluation methodology for shared workspace groupware. ACM conference on Computer supported cooperative work, 96-105.
- Muller, Michael J., Matheson, L., Page, C., Gallup, R. (1998) Methods & tools: participatory heuristic evaluation. interactions, Volume 5 Issue 5, 13-18.

computer operator, 1961 (IBM 7030)



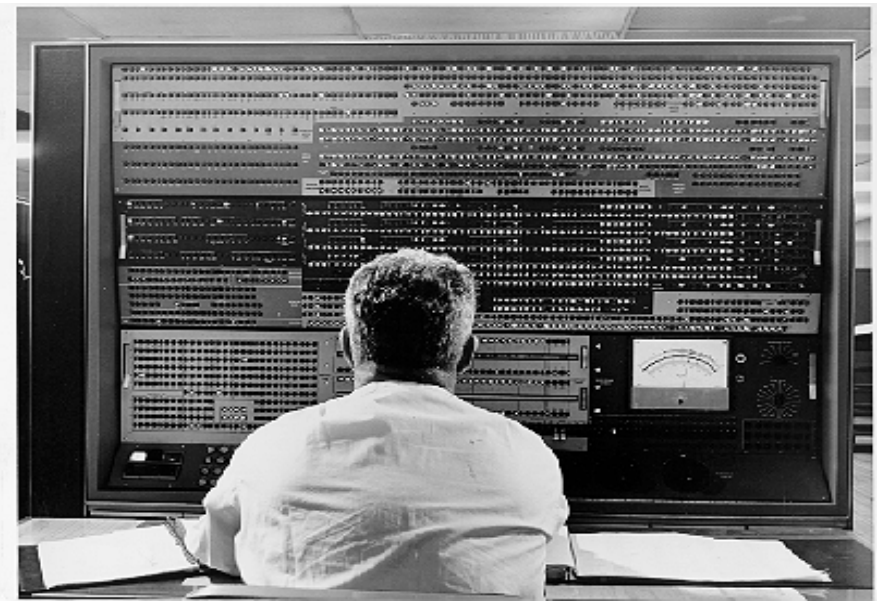
http://archive.computerhistory.org/stretch

Bloch (1959): "The Engineering Design of the Stretch Computer", Proc. ACM Eastern Joint Computer Conference, 48-58



a medical doctor, 2005

a computer scientist, 1961



+ Phases of Heuristic Evaluation

1) Pre-evaluation training
   - give evaluators needed domain knowledge and information on the scenario
2) Evaluation
   - individuals evaluate and then aggregate results
3) Severity rating
   - determine how severe each problem is (priority)
     - can do this first individually and then as a group
4) Debriefing
   - discuss the outcome with design team

+ How to Perform Evaluation

- At least two passes for each evaluator
  - first to get feel for flow and scope of system
  - second to focus on specific elements
- If system is walk-up-and-use or evaluators are domain experts, no assistance needed
  - otherwise might supply evaluators with scenarios
- Each evaluator produces list of problems
  - explain why with reference to heuristic or other information
  - be specific and list each problem separately

+ Severity Ratings

0 - don't agree that this is a usability problem
1 - cosmetic problem
2 - minor usability problem
3 - major usability problem; important to fix
4 - usability catastrophe; imperative to fix

Results of Using HE

- Discount: benefit-cost ratio of 48 [Nielsen94]
  - cost was $10,500 for benefit of $500,000
  - value of each problem ~15K (Nielsen & Landauer)
  - how might we calculate this value?
    - in-house -> productivity; open market -> sales
- Correlation between severity & finding w/ HE
- Single evaluator achieves poor results
  - only finds 35% of usability problems
  - 3-5 evaluators find ~ 75% of usability problems
  - why not more evaluators???? 10? 20?
    - adding evaluators costs more & won't find more probs

+ Group interviews

- Also known as 'focus groups'
- Typically 3-10 participants
- Provide a diverse range of opinions
- Need to be managed to:
  - ensure everyone contributes
  - discussion isn't dominated by one person
  - the agenda of topics is covered

+ Analyzing interview data

- Depends on the type of interview
- Structured interviews can be analyzed like questionnaires
- Unstructured interviews generate data like that from participant observation
- It is best to analyze unstructured interviews as soon as possible to identify topics and themes from the data

+ Questionnaires

- Questions can be closed or open
- Closed questions are easiest to analyze, and may be done by computer
- Can be administered to large populations
- email can be used for dissemination
- Advantage of electronic questionnaires is that data goes into a data base & is easy to analyze
- Sampling can be a problem when the size of a population is unknown as is common online

+ Questionnaire style

- Varies according to goal
- Questionnaire format can include:
  - 'yes', 'no' checkboxes
  - checkboxes that offer more options
  - Likert rating scales
  - semantic scales
  - open-ended responses
- Likert scales have a range of points
- 3, 5, 7 point scales are common
- Debate in team about which suites best!

+ Developing a questionnaire

- Provide a clear statement of purpose & guarantee participants anonymity
- Plan questions - if developing a web-based questionnaire, design off-line first
- Decide on whether phrases will all be positive, all negative or mixed
- Pilot test questions - are they clear, is there sufficient space for responses
- Decide how data will be analyzed & consult a statistician if necessary

+ Encouraging a good response

- Make sure purpose of study is clear
- Provide anonymity
- Ensure questionnaire is well designed
- Offer a short version for those who do not have time to complete a long questionnaire
- Follow-up with phone calls
- Provide an incentive
- 40% response rate is high, 20% is often acceptable, mostly you get back 10 %!

Questionnaire data analysis & presentation

- Present results clearly - tables may help
- Simple statistics can say a lot, e.g., mean, median, mode, standard deviation
- Percentages are useful but give population size
- Bar graphs show categorical data well
- More advanced statistics can be used if needed (-> SPSS Data Analysis)

SUMI

- http://www.ucc.ie/hfrg/questionnaires/sumi

MUMMS

http://www.ucc.ie/hfrg/questionnaires/mumms

QUIS™

http://www.lap.umd.edu/QUIS

http://www.ucc.ie/hfrg/resources/qfaq1.html#whatisaquestionnaire

- Focus on ease of learning
- Designer presents an aspect of the design & usage scenarios
- One of more experts walk through the design prototype with the scenario
- Expert is told the assumptions about user population, context of use, task details
- Experts are guided by 3 questions

+ The 3 questions

- Will the correct action be sufficiently evident to the user?
- Will the user notice that the correct action is available?
- Will the user associate and interpret the response from the action correctly?

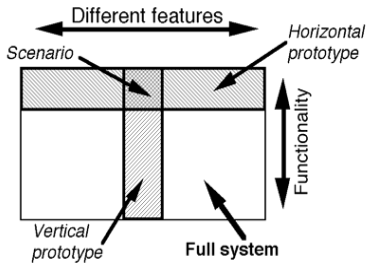
As the experts work through the scenario they note problems

+ Pluralistic walkthrough

- Variation on the cognitive walkthrough theme
- Performed by a carefully managed team
- The panel of experts begins by working separately
- Then there is managed discussion that leads to agreed decisions
- The approach lends itself well to participatory design

- Questionnaire to test reactions with friends
- http://www.acm.org/~perlman/question.html
- http://www.ifsm.umbc.edu/djenni1/osg/
- Develop heuristics to evaluate usability and sociability aspects

- Cut down complexity by eliminating parts of full system
- Small, cheap to design and implement
  - Paper mock-ups



Bodker, S. (2000), Scenarios in user-centred design:
Setting the stage for reflection and action. *Interacting with Computers*, 13, 1, 61-75.

Cognitive Walkthrough

- Evaluation by an expert, who goes through a set task while imitating user performance

 Preparation: We need four things

1. User description including level of experience with computers and any assumptions made by the designer
2. System description including operations and performance (e.g. paper design)
3. Task description specifying the task that the expert has to carry out from users point of view
4. Action sequence describing the system display and the user actions needed to complete the given task. One system display and one user action together are one step.

Process of a Cognitive Walkthrough

- Prototype, user description, system description, task description and action list has to be prepared by the designer/developer
- Designer gives these document to an expert.
- The expert reads the descriptions.
- The expert carries out the task by following the action list.
- The expert asks the following questions with EACH step of the action list:

Questions Cognitive Walkthrough

1. Is the next goal clear at this stage?
2. Is the appropriate action obvious?
3. Is it clear that this action leads to the goal?
4. What problems are there in performing the action?

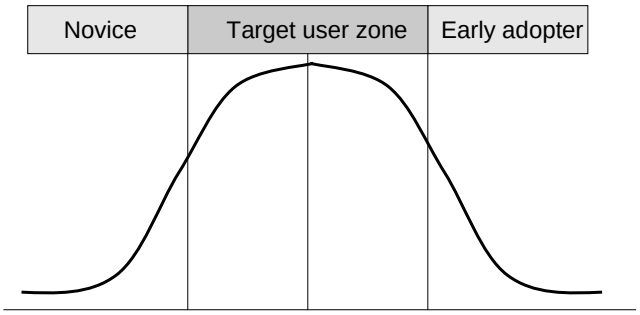
Four Questions Asked for Each Action while achieving the goal

- Will the user try to achieve the right effect?
- Will the user notice the correct action is available?
- Will the user associate the correct action with the effect that the user is trying to achieve?
- If the correct action is performed, will the user see that progress is being made toward solution of the task?

- GOMS
  - **G**oal: User's goal, describe what the user want to achieve
  - **O**perators: Basic actions that the user must perform in order to use the system
  - **M**ethods: Several ways in which a goal can be split into sub-goals
  - **S**election: choose from methods according to certain rule

Reitman, J. & Olson, G. M. (1990), The growth of cognitive modeling in human-computer interaction since GOMS. *Human-Computer Interaction*, 5, 221-265.

- Two main elements:
  - Cost estimation
  - Benefit estimation

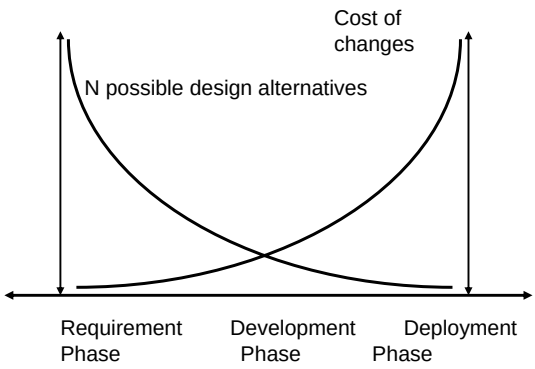


New user
-Tentative
-May need proactive help

Early adopter
- fast
- aggressive
- needs little /no hel
- values fast, efficie
  user experience

Donahue (2002)

Creeping costs



Creeping costs

- Total maintenance costs $20-30billion/year
- Backlog maintenance minimum 167% of this (... 80% of all problems are identified during maintenance, and it is estimated that 80% of these could be saved with a thorough UCD!

  - Martin & McClune, 1998)

- 2 parameters are important:
  - Cost/benefit ratio
    - Relationship between projected use of resources and post-implementation/post-sale value of the investment (and savings during development)
  - Payback period
    - The amount of time it will take before the product starts to outweigh the cost of the investment

- Cost/benefit ratio
  - Relationship between projected use of resources and post-implementation/post-sale value of the investment (and savings during development)
- Cost
  - Team of 4, 2 hours @ $60 = $480
- Saving
  - 4 person days = 32 hours @ $60/hour = $1,920
- Cost/benefit ratio = 1:4

- First Stage: skepticism
- Second Stage: curiosity
- Third Stage: Acceptance Stage
  - DUE sporadically used
  - Systematically used
- Fourth Stage: Partnership Stage
  - lab founded
  - Usability permeates lifecycle

- Heute 3 Aufgaben:
- 1) Entwickeln Sie einen Fragebogen, um Information über Ihre Applikation zu gewinnen
- 2) Entwickeln Sie ein Modell zur Bestimmung der Cost-Benefit Ratio Ihrer Applikation
- 3) Diskutieren Sie Mehrwerte Ihrer Applikation durch den möglichen Einsatz multimodaler Interfaces!

Los gehts!