emex

EMERGING MEDIA EXPLORATION

Evaluation of Interactive Systems

Jaakko Hakulinen, Päivi Majaranta

EMEX workshop 10.10.2019

emex

Evaluation approaches

- Exploratory vs. clear goals
- Usability studies
  - a part of the iterative development
  - must be concluded before use experience evaluation phase starts
- Objective vs. subjective data
- Laboratory vs. in-the-wild evaluations
- Booking participantes vs. guerilla evaluations
- Comparative studies
  - Within subject – between subjects
- Real system – Partly simulated - Wizard of Oz – Paper Prototype

emex

What is the Goal of the Evaluation

- Support development
  - Help ideation in early phases of the project
  - Identify usability problems
  - Identify technical performance and technical issues
  - Choose the best parameters/solutions
  - Better understand user needs and attitudes
- Validate design
- Support selection between systems
- Generate marketing material
- Generate scientific knowledge
  - Generalizable understanding

emex

Participants

- In usability and UX evaluations, the selection of participants is important
  - Number of participants (usability evaluations with 5 people are efficient, statistical comparisons of UX measures require over 10 people, often much more)
  - Background: age, gender, native tongue…
  - Domain understanding: professional systems often require domain experts
- In "regular" testing you arrange a time for each participant before hand, in guerrilla testing you approach people and ask if they can spend a few minutes with your.
  - guerrilla testing must be fast while regular testing can take up to one hour

emex

Type of Result Data

	Objective	Subjective
Quantitative	Performance metrics - Task completion times - Number of turns - … Behavioural data Bio-signals - Heart rate - Electric skin conductivity - …	Answers to binary and scale-based questions in questionnaires and interviews Numerically analysed data from interviews
Qualitative		Interviews, both audio and transcribed text Video recordings Participants' drawings …

emex

Type of Result Data

- Different types of data provide different information
- Quantitative data, i.e., numbers are useful when you want to compare things and have hard numbers
- Qualitative data is valuable in all phases but most important early on
- An interview is a good tool in many phases
  - early you get good understanding of how people perceive your concept and also why they do so
  - later in usability testing you get better understanding of reasons of the problems found
  - in late phases, you can still get valuable information on the content/product was received
- In most cases, interview doesn't need to be transcribed or analysed in detail but there is methodology for this when interviews are main data in scientific research.

emex

Evaluation Procedures

- Evaluations can
  - have a strict procedure
    - Participants are given tasks to complete
    - Feedback is collected at specified times with specified tools
  - have an open procedure
    - Participants are allowed to freely use or explore the system
    - Feedback is collected at appropriate times
  - be done as part of the real use
    - Participants are real users doing their real life tasks
    - Feedback collection may be integrated into the system
- In all cases, if there is a system with some implemented functionality, the system can log data of how it is used.

emex

Think-aloud

- By asking the participants to speak aloud what they are doing and thinking while using your prototype, you can learn more than just by observing their behaviour.

- Some people are better at this than others, for some it is not natural at all.

  - you can remind people about think-aloud but don't force people

- Think-aloud does have some effect of participant performance so it should not be used if, e.g., performance times are measured and compared.

- In some cases, think-aloud is not possible, e.g., if the system is a voice user interface.

emex

Wizard of Oz (WoZ)

- One (or more) part of the system is replaced by a human (sometimes called human-in-the-loop)

- For example: human replaces speech recognizer

- The human needs an efficient interface to minimize the delays

- Good instructions and training must also be given so that the system behaves in consistent and desired way

- Suitable method when relevant technology is not available yet or not yet good enough

emex

Location of Evaluation

- Laboratory is a controlled environment
  - Effects of external conditions minimized, focus on controlled variables
  - Some effects of real use context can be simulated
    - Noise via speakers
    - Movement using a treadmill
- Real use context is harder to control and data collection can be challenging, but conditions are more realistic
  - Real environmental conditions (noise, temperature, lighting…)
  - Real social situation
  - Context can affect participants' state of mind and associations they make
- In-the-wild (usually) refers to the real use context with minimal control to of environment and selection of participants from people in the environment
- Studies have shown that significant portion of usability issues can be found only in real use context:
  - Start in the laboratory but go to field soon

emex

Ethical Points

- Participants are voluntary
    - they must be allowed to decline request to participate
    - they must be allowed to stop and leave the evaluation at any point, without giving their reasons
- Respect participants' privacy, only collect the personal information you really need, ask permission to take photos, video and audio recordings
- Do not cause unnecessary stress to the participants
    - evaluations should be reasonably short
    - explain what will happen if there is anything that may be considered uncomfortable by somebody

emex