Design of Interactive Systems (DIS)

Lecture 7: Techniques for designing interactive systems - Evaluation

Dr. Kalpana Shankhwar, *kalpana@iiitd.ac.in*
Assistant Professor
Department of Human Centered Design,
IIIT Delhi

Part II Techniques for designing interactive systems

- **Chapter 7**: Understanding
- **Chapter 8**: Envisionment
- **Chapter 9**: Design
- **Chapter 10**: Evaluation
- **Chapter 11**: Task Analysis
- **Chapter 12**: Visual Interface Design
- **Chapter 12**: Multimodal Interface Design

Evaluation

- **Forth main process** of interaction systems design
- **Reviewing**, **testing** a design idea, piece of software, product or service
- Whether the systems meet some **criteria**:
  - System is *learnable*, *effective* and *accommodating*
  - Is this enough?
- Designers are concerned not just with surface features such as the meaningfulness of icons, but also with whether the **system is fit for its purpose, enjoyable, engaging** and so on.

Aim

- Study the evaluation techniques designed for use **with** and **without users**
- Understand **expert-based evaluation** methods
- Understand **participant-based evaluation** methods
- Apply the techniques in **appropriate contexts**.

Challenges in Evaluation

- Evaluation of **different types of system**, or evaluation in **different contexts**, may offer particular challenges.
- Evaluation is closely tied to *understanding*, *design* and *envisionment*
- There are issues concerning **who is involved** in the evaluation?

Evaluation Types

- There are two main types of evaluation
- **Expert-based methods**
  - reviewing some form of envisioned version of a design by **usability expert** or **interaction designer**
- **Participant methods**
  - Involves recruiting people (**end-users**) to use an envisioned version of a system
  - The characteristics of the target population should be captured through personas

Evaluation

- Evaluation occurs **throughout interaction design** process
- At **different stages**, **different methods** will be more or less effective
- **Form of envisionment** of the future systems is also critical to what can be evaluated
- Obtaining feedback to inform early design concepts – to check novelty through quick prototyping
- Deciding between different design options

Checking for Usability Problems

- Testing will identify potential problems once a stable version of the technology is available
  - **Vertical** or **horizontal prototype**
- Check that interaction is '**user-friendly**' just before development is completed
- It is also termed **formative evaluation** because the results help to form - or shape - the design
- Government departments often require suppliers to conform to **accessibility standards** and **health** and **safety legislation**

Expert Evaluation

- There is **no** substitute for getting real people to use your design, but expert evaluation is effective, particularly early in the design process
- Experts will pick up common problems based on this experience, and will identify factors that might otherwise interfere with an evaluation by non-experts
- Expert will **walk through** representative tasks or scenarios of use
- Additionally they may *adopt one of the personas*.

Expert Evaluation: Heuristic evaluation

- It refers to a number of methods in which a person trained in HCI and interaction design examines a proposed design to see how it measures up against a list of principles, guidelines or ‘heuristics’ for good design
- design principles
  - or heuristics

1. Visibility
2 Consistency
3 Familiarity
4 Affordance
5 Navigation
6 Control
7 Feedback
8 Recovery
9 Constraints
10 Flexibility
11 Style
12 Conviviality

Expert Evaluation: Discount usability engineering

- If time is very short, a quick review of the design against the following triad can produce reasonably useful results
- usability principles:
  - *learnability* (1-4)
  - *effectiveness* (5-9)
  - *Accommodation (10-12)*

1. Visibility
2 Consistency
3 Familiarity
4 Affordance
5 Navigation
6 Control
7 Feedback
8 Recovery
9 Constraints
10 Flexibility
11 Style
12 Conviviality

- **You should not evaluate your own designs.**

Expert Evaluation: Cognitive walkthrough

It is a rigorous paper-based technique for checking through the detailed design and logic of steps in an interaction

It is a powerful and well-established theory-based approach.

Inputs to the process are:

- An understanding of the people who are **expected to use** the system
- A set of concrete scenarios representing both (a) **very common** and (b) **uncommon but critical sequences of activities**
- A complete **description of the interface** to the system - this should comprise both a representation of how the interface is presented, e.g. **screen designs**, and the **correct sequence of actions** for achieving the scenario tasks, usually as a hierarchical task analysis (HTA).

Expert Evaluation: Cognitive walkthrough

Having gathered these materials together, the analyst asks the following four questions for each individual step in the interaction:

- Will the people using the system try to achieve the right effect?
- Will they notice that the correct action is available?
- Will they associate the correct action with the effect that they are trying to achieve?
- If the correct action is performed, will people see that progress is being made towards the goal of their activity?

Expert Evaluation: Cognitive walkthrough

- What if any of answers in negative?
  - usability problem has been identified and is recorded
  - redesign suggestions are not made at this point
- If the walkthrough is being used as originally devised, this process is carried out as a **group exercise** by analysts and designers together
- The analysts step through usage scenarios and the design team are required to explain how the **user would identify, carry out and monitor the correct sequence of actions**

Cut-down Versions of Cognitive walkthrough

- **cognitive jogthrough**
  - video records are made of walkthrough meetings
  - annotated to indicate significant items of interest
  - design suggestions are permitted
  - low-level actions are aggregated wherever possible
- **streamlined cognitive walkthrough**
  - Will people know what to do at each step?
  - If people do the right thing, will they know that they did the right thing, and are making progress towards their goal?

Participant-based Evaluation

- Various degrees of cooperation with people
- **Cooperative evaluation**
  - a means of maximizing the data gathered from a simple testing session
  - participants are **not passive subjects** but work as **co-evaluators**
  - a reliable but economical technique in diverse applications
- Participatory heuristic evaluation
- Co-discovery
- Controlled experiments

- Cooperative evaluation guidelines
- Prepare questions **during** and **after the evaluation**

Table 10.1 Guidelines for cooperative evaluation

Step	Notes
1 Using the scenarios prepared earlier, write a draft list of tasks.	Tasks must be realistic, doable with the software, and explore the system thoroughly.
2 Try out the tasks and estimate how long they will take a participant to complete.	Allow 50 per cent longer than the total task time for each test session.
3 Prepare a task sheet for the participants.	Be specific and explain the tasks so that anyone can understand
4 Get ready for the test session.	Have the prototype ready in a suitable environment with a list of prompt questions, notebook and pens ready. A video or audio recorder would be very useful here.
5 Tell the participants that it is the system that is under test, not them; explain and introduce the tasks.	Participants should work individually – you will not be able to monitor more than one participant at once. Start recording if equipment is available.
6 Participants start the tasks. Have them give you running commentary on what they are doing, why they are doing it and difficulties or uncertainties they encounter.	Take notes of where participants find problems or do something unexpected, and their comments. Do this even if you are recording the session. You may need to help if participants are stuck or have them move to the next task.
7 Encourage participants to keep talking.	Some useful prompt questions are provided below.
8 When the participants have finished, interview them briefly about the usability of the prototype and the session itself. Thank them.	Some useful questions are provided below. If you have a large number of participants, a simple questionnaire may be helpful.
9 Write up your notes as soon as possible and incorporate into a usability report.	

Sample questions *during* the evaluation:
- What do you want to do?
- What were you expecting to happen?
- What is the system telling you?
- Why has the system done that?
- What are you doing now?

Sample questions *after* the session:
- What was the best/worst thing about the prototype?
- What most needs changing?
- How easy were the tasks?
- How realistic were the tasks?
- Did giving a commentary distract you?

Participant-based Evaluation

- Cooperative evaluation
- **Participatory heuristic evaluation**
  - it extends the power of heuristic evaluation without adding greatly to the effort required
  - the participants are involved as ‘work-domain experts’ alongside usability experts and must be briefed about what is required
- Co-discovery
- Controlled experiments

Participant-based Evaluation

- Cooperative evaluation
- Participatory heuristic evaluation
- **Co-discovery**
  - A naturalistic, informal technique that is particularly good for capturing first impressions
  - best used in the later stages of design
  - having participants explore new technology in pairs
  - Better to have people in pair who know eachother quite well
- Controlled experiments

Participant-based Evaluation

- Cooperative evaluation
- Participatory heuristic evaluation
- Co-discovery
- **Controlled experiments**
  - appropriate where the designer is interested in particular features of a design
  - experiment needs to be carefully designed and run
  - For example, comparing two different designs of a website or two different ways of selecting a function on an App.

Evaluation in Practice

The main steps in undertaking a simple but effective evaluation project are:

1. Establish the **aims of the evaluation**, the **intended participants** in the evaluation, the **context of use** and the **state of the technology**; obtain or construct scenarios illustrating **how the application will be used**.
2. Select evaluation methods. These should be a combination of expert-based review methods and participant methods.
3. Carry out expert review.
4. Plan participant testing; use the results of the expert review to help focus this.
5. Recruit people and organize testing venue and equipment.
6. Carry out the evaluation.
7. Analyse results, document and report back to designers.

Aims of the evaluation

- Deciding the aims for evaluation helps to determine the type of data required. For example, in the evaluation of the early concept for a virtual training environment the aims were to investigate:
  - Do the trainers understand and welcome the basic idea of the virtual training environment?
  - Would they use it to extend or replace existing training courses?
  - How close to reality should the virtual environment be?
  - What features are required to support record keeping and administration?

Aims of the evaluation

- The data we were interested in at this stage was largely **qualitative** (non-numerical), so appropriate data gathering methods were **interviews** and **discussions with the trainers**.
- If the aim of the evaluation is the comparison of two different **evaluation designs** then much more focused questions will be required and the data gathered will be more **quantitative**.
- For example, the questions were:
  - Is it quicker to reach a particular room in the virtual environment using mouse, joystick?
  - Is it easier to open a virtual door by clicking on the handle or selecting the “open” icon from a tool palette?

Metrics and measures

- What is to be measured and how? Table 10.3 shows some common usability metrics
- Such metrics are helpful in evaluating many types of applications, from small mobile communication devices to office systems.

Table 10.3 Common usability metrics

Usability objective	Effectiveness measures	Efficiency measures	Satisfaction measures
Overall usability	• Percentage of tasks successfully completed • Percentage of users successfully completing tasks	• Time to complete a task • Time spent on non-productive actions	• Rating scale for satisfaction • Frequency of use if this is voluntary (after system is implemented)
Meets needs of trained or experienced users	• Percentage of advanced tasks completed • Percentage of relevant functions used	• Time taken to complete tasks relative to minimum realistic time	• Rating scale for satisfaction with advanced features
Meets needs for walk up and use	• Percentage of tasks completed successfully at first attempt	• Time taken on first attempt to complete task • Time spent on help functions	• Rate of voluntary use (after system is implemented)
Meets needs for infrequent or intermittent use	• Percentage of tasks completed successfully after a specified period of non-use	• Time spent re-learning functionsNumber of persistent errors	• Frequency of reuse (after system is implemented)
Learnability	• Number of functions learned • Percentage of users who manage to learn to a pre-specified criterion	• Time spent on help functions • Time to learn to criterion	• Rating scale for ease of learning

People

- The most **important people** in evaluation are the people **who will use the system**.
- Analysis work should have identified the characteristics of these people, and represented these in the form of **personas**.
- **Nielsen's** recommended sample of **3-5 participants** has been accepted wisdom in usability practice for over a decade.
- Small number only helps for a **homogeneous** group
- Students are often readily available, but they are only representative of a **particular segment** of the population. If you have the resources, **payment** can help recruitment.

People

- Inevitably, your sample will be biased towards **cooperative people** with some sort of interest in **technology**, so bear this in mind when interpreting your results.
- For receiving brutally **honest reaction**, one of your colleagues, a friend, your mother or anyone you trust can be considered
- Recommended method for basic testing requires an evaluator to sit with each user and engage with them as they carry out the test tasks.
- you should provide help if the participant is becoming uncomfortable, or completely stuck.

The test plan and task specification

A plan should be drawn up to guide the evaluation. The plan specifies:

- **Aims** of the test session
- Practical details, including **where** and **when** it will be conducted, **how long** each session will last, the **specification** of equipment and materials for testing and data collection, and any **technical support** that may be necessary
- **Numbers** and **types** of participant
- Tasks to be performed, with a definition of successful completion. This section also specifies **what data** should be collected and **how** it will be analysed.

Reporting usability evaluation results to design team

- Even if you are both designer and evaluator, you need an organized list of findings so that you can prioritize redesign work.
- If you are reporting back to a design/ development team
  - it is crucial that they can see immediately what the problem is,
  - how significant its consequences are,
  - and ideally what needs to be done to fix it.

Evaluation: further issues

Evaluation without being there

- With the arrival of Internet connectivity, people can participate in evaluations without being physically present.
- If the application itself is Web-based, or can be installed remotely, instructions can be supplied so that users can run test tasks and fill in and return questionnaires in soft or hard copy.

Evaluation: further issues

Physical and physiological measures

- Eye-movement tracking (or ‘eye tracking’) can show participants’ changing focus on different areas of the screen.
- This can indicate which features of a user interface have attracted attention, and in which order,
- Physiological techniques in evaluation rely on the fact that all our emotions - anxiety, pleasure, apprehension, delight, surprise and so on - generate physiological changes.

Evaluation: further issues

Physical and physiological measures

- The most common measures are of changes in heart rate, the rate of respiration, skin temperature, blood volume pulse and galvanic skin response
- Sensors can be attached to the participant's body and linked to software which converts the results to numerical and graphical formats for analysis.

Evaluating presence

- Designers of virtual reality - and some multimedia - applications are often concerned with the sense of presence, of being ‘there’ in the virtual environment rather than ‘here’ in the room where the technology is being used.
- A strong sense of presence is thought to be crucial for such applications as games, those designed to treat phobias, to allow people to ‘visit’ real places they may never see otherwise,

Evaluation at home

- People at home are much less of a '**captive audience**' for the evaluator than those at work.
- They are also likely to be more concerned about **protecting their privacy** and generally unwilling to spend their **valuable leisure time** in helping you with your usability evaluation.
- So it is important that data gathering techniques are **interesting** and **stimulating** for users, and make as **little demand on time and effort** as possible.

Class activity

Class activity

This video demonstrates a virtual reality application for enhancing the online shopping experience with haptic feedback to touch the objects and experience the texture. In order to evaluate its **usability** and the **sense of presence,** prepare a questionnaire for cooperative evaluation using the guidelines mentioned in Table 10.1.