Human Computer Interaction (INSY4112)

CHAPTER 7

Evaluation Techniques and Universal Design

EVALUATION TECHNIQUES AND UNIVERSAL DESIGN

- Evaluation Techniques
  - What is evaluation?
  - Goals of evaluation
  - Choosing an evaluation method

- Universal Design
  - Introduction
  - Universal design principles
  - Multi-modal interaction
  - Designing for diversity

Objectives

- To identify evaluation techniques and methods
- To describe universal design principles
- To explain Multi-modal interaction
- To describe the goal of evaluation

Evaluation Techniques

➢ Evaluation

- It tests usability, functionality and acceptability of an interactive system
- Evaluation occurs in laboratory, field and/or in collaboration with users
- It evaluates both design and implementation
- It should be considered at all stages in the design life cycle

Evaluation Techniques

➢ Evaluation

- Evaluation may take place:
  - ✓ in the laboratory
  - ✓ in the field.
- Some approaches are based on expert evaluation:
  - ✓ analytic methods
  - ✓ review methods
  - ✓ model-based methods.

Evaluation Techniques

➢ Evaluation

- Some approaches involve users:
  - ✓ experimental methods
  - ✓ observational methods
  - ✓ query methods.
- An evaluation method must be chosen carefully and must be suitable for the job.

Evaluation Techniques

➢ Goals of Evaluation

There are three goals:

1. assess extent of system functionality

2. assess effect of interface on user

3. identify specific problems

Evaluation Techniques

➢ Evaluating Designs

Cognitive Walkthrough

Heuristic Evaluation

Review-based evaluation

Evaluation Techniques

➢ Cognitive Walkthrough

- ❖ Walkthroughs require a detailed review of a sequence of actions.
- ❖ Proposed by Polson *et al.*
  - Evaluates design on how well it supports user in learning task
  - Usually performed by expert in cognitive psychology
  - Expert 'walks though' design to identify potential problems using psychological principles
  - Forms used to guide analysis

Evaluation Techniques

➢ Cognitive Walkthrough

❖ for each action, the evaluators try to answer the following four questions for each step in the action sequence:

1. Is the effect of the action the same as the user's goal at that point?
2. Will users see that the action is available?
3. Once users have found the correct action, will they know it is the one they need?
4. After the action is taken, will users understand the feedback they get? This is the completion of the execution–evaluation interaction cycle

Evaluation Techniques

➢ Cognitive Walkthrough (ctd…)

- ❖ For each task walkthrough considers
  - – what impact will interaction have on user?
  - – what cognitive processes are required?
  - – what learning problems may occur?
- ❖ Analysis focuses on goals and knowledge: does the design lead the user to generate the correct goals?

Evaluation Techniques

➢ Heuristic Evaluation

- Proposed by Nielsen and Molich.
- usability criteria (heuristics) are identified
- design examined by experts to see if these are violated
- Heuristic evaluation `debugs' design.

Evaluation Techniques

➢ Heuristic Evaluation

- Nielsen's ten heuristics are:

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standard
5. Error prevention

Evaluation Techniques

➢ Heuristic Evaluation

- Nielsen's ten heuristics are:

6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist
9. Help users recognize, diagnose and recover from errors
10. Help and documentation

Evaluation Techniques

➢ Review-based evaluation

- ❖ Results from the literature used to support or refute parts of design.
- ❖ Care needed to ensure results are transferable to new design.

➢ Model-based evaluation

- ❖ Cognitive models used to filter design options

  e.g. GOMS prediction of user performance.

- ❖ Design rationale can also provide useful evaluation information

Evaluation Techniques

➢ Evaluating through user Participation

❖ There are different approaches to evaluation through user participation. These include

i. empirical or experimental methods,

ii. observational methods,

iii. query techniques, and

iv. methods that use physiological monitoring

Evaluation Techniques

➢ There are two distinct evaluation styles

a) Laboratory studies

- Advantages:
  - specialist equipment available
  - uninterrupted environment
- Disadvantages:
  - lack of context
  - difficult to observe several users cooperating
- Appropriate
  - if system location is dangerous or impractical for constrained single user systems to allow controlled manipulation of use

b. Field Studies

- Advantages:
  - natural environment
  - context retained (though observation may alter it)
  - longitudinal studies possible
- Disadvantages:
  - distractions
  - noise
- Appropriate
  - where context is crucial for longitudinal studies

Evaluation Techniques

➢ Evaluating Implementations

- ❖Requires an artefact:
  - simulation,
  - prototype,
  - full implementation

Evaluation Techniques

➢ Experimental evaluation

- ❖ controlled evaluation of specific aspects of interactive behaviour
- ❖ evaluator chooses hypothesis to be tested
- ❖ a number of experimental conditions are considered which differ only in the value of some controlled variable.
- ❖ changes in behavioural measure are attributed to different conditions

Evaluation Techniques

➢ Experimental factors

- ❖ Subjects
  - – who – representative, sufficient sample
- ❖ Variables
  - – things to modify and measure
- ❖ Hypothesis
  - – what you'd like to show
- ❖ Experimental design
  - – how you are going to do it

Evaluation Techniques

➢ Variables

- Independent variable (IV)
  - ✓ characteristic changed to produce different conditions
  - ✓ e.g. interface style, number of menu items

- Dependent variable (DV)
  - ✓ characteristics measured in the experiment
  - ✓ e.g. time taken, number of errors.

Evaluation Techniques

➢ Hypothesis

- prediction of outcome
  - framed in terms of IV and DV

    e.g. “error rate will increase as font size decreases”
- null hypothesis:
  - states no difference between conditions
  - aim is to disprove this

    e.g. null hyp. = “no change with font size”

Evaluation Techniques

➢ Experimental design

- ❖ Within groups design
  - each subject performs experiment under each condition.
  - transfer of learning possible
  - less costly and less likely to suffer from user variation.
- ❖ Between groups design
  - each subject performs under only one condition
  - no transfer of learning
  - more users required
  - variation can bias results.

Evaluation Techniques

➢ Analysis of data

- Before you start to do any statistics:
  - look at data
  - save original data
- Choice of statistical technique depends on
  - type of data
  - information required
- Type of data
  - discrete - finite number of values
  - continuous - any value

Evaluation Techniques

➢ Analysis - types of test

- parametric
  - assume normal distribution
  - robust
  - powerful
- non-parametric
  - do not assume normal distribution
  - less powerful
  - more reliable
- contingency table
  - classify data by discrete attributes
  - count number of data items in each group

Evaluation Techniques

➢ Analysis of data (cont.)

- What information is required?
  - is there a difference?
  - how big is the difference?
  - how accurate is the estimate?
- Parametric and non-parametric tests mainly address first of these

Evaluation Techniques

➢ Experimental studies on groups

- More difficult than single-user experiments
- Problems with:
  - subject groups
  - choice of task
  - data gathering
  - analysis

Evaluation Techniques

➢ Subject groups

- ❖ Larger number of subjects ⇒ more expensive
- ❖ Longer time to `settle down'… even more variation!
- ❖ Difficult to timetable
- ❖ So … often only three or four groups

Evaluation Techniques

➢ The task

- ❖ must encourage cooperation
- ❖ perhaps involve multiple channels
- ❖ options:
  - creative task — e.g. '*write a short report on …*'
  - decision games — e.g. desert survival task
  - control task — e.g. ARKola bottling plant

Evaluation Techniques

➢ Data gathering

- several video cameras + direct logging of application
- problems:
  - synchronisation
  - sheer volume!
- one solution:
  - record from each perspective

Evaluation Techniques

➢ Analysis

- N.B. vast variation between groups
- solutions:
  - within groups experiments
  - micro-analysis (e.g., gaps in speech)
  - anecdotal and qualitative analysis
- look at interactions between group and media
- controlled experiments may `waste' resources!

Evaluation Techniques

➢ Field studies

- Experiments dominated by group formation
- Field studies more realistic:
  - ✓ *distributed cognition* ⇒ work studied in context
  - ✓ real action is *situated action*
  - ✓ physical and social environment both crucial
- Contrast:
  - ✓ psychology – controlled experiment
  - ✓ sociology and anthropology – open study and rich data

Evaluation Techniques

➢ Observational Methods

- ❖ Think Aloud
- ❖ Cooperative evaluation
- ❖ Protocol analysis
- ❖ Automated analysis
- ❖ Post-task walkthroughs

Evaluation Techniques

➢ Think Aloud

- User observed performing task
- User asked to describe what he is doing and why, what he thinks is happening etc.
- Advantages
  - simplicity - requires little expertise
  - can provide useful insight
  - can show how system is actually use
- Disadvantages
  - subjective
  - selective
  - act of describing may alter task performance

Evaluation Techniques

➢ Cooperative evaluation

- Variation on think aloud
- User collaborates in evaluation
- Both user and evaluator can ask each other questions throughout
- Additional advantages
  - less constrained and easier to use
  - user is encouraged to criticize system
  - clarification possible

Evaluation Techniques

➢ Protocol analysis

- ❖ paper and pencil – cheap, limited to writing speed
- ❖ audio – good for think aloud, difficult to match with other protocols
- ❖ video – accurate and realistic, needs special equipment, obtrusive
- ❖ computer logging – automatic and unobtrusive, large amounts of data difficult to analyze

Evaluation Techniques

➢ Protocol analysis

- ❖ user notebooks – coarse and subjective, useful insights, good for longitudinal studies
- ❖ Mixed use in practice.
- ❖ audio/video transcription difficult and requires skill.
- ❖ Some automatic support tools available

Evaluation Techniques

➢ Automated Analysis – Evaluation

- Workplace project
- Post task walkthrough
  - user reacts on action after the event
  - used to fill in intention
- Advantages
  - analyst has time to focus on relevant incidents
  - avoid excessive interruption of task
- Disadvantages
  - lack of freshness
  - may be post-hoc interpretation of events

Evaluation Techniques

➢ Post-task walkthroughs

- ❖ transcript played back to participant for comment
  - – immediately → fresh in mind
  - – delayed → evaluator has time to identify questions
- ❖ useful to identify reasons for actions and alternatives considered
- ❖ necessary in cases where think aloud is not possible

Evaluation Techniques

➢ Query Techniques

- Interviews
- Questionnaires

Evaluation Techniques

➢ Interviews

- analyst questions user on one-to -one basis

  usually based on prepared questions
- informal, subjective and relatively cheap
- Advantages
  - can be varied to suit context
  - issues can be explored more fully
  - can elicit user views and identify unanticipated problems
- Disadvantages
  - very subjective
  - time consuming

Evaluation Techniques

➢ Questionnaires

- Set of fixed questions given to users
- Advantages
  - quick and reaches large user group
  - can be analyzed more rigorously
- Disadvantages
  - less flexible
  - less probing

Evaluation Techniques

➢ Questionnaires (ctd)

- Need careful design
  - what information is required?
  - how are answers to be analyzed?
- Styles of question
  - general
  - open-ended
  - scalar
  - multi-choice
  - ranked

Evaluation Techniques

- **Physiological methods**

- Eye tracking
- Physiological measurement

Evaluation Techniques

➢ Eye tracking

- ❖ head or desk mounted equipment tracks the position of the eye
- ❖ eye movement reflects the amount of cognitive processing a display requires
- ❖ measurements include
  - fixations: eye maintains stable position. Number and duration indicate level of difficulty with display
  - saccades: rapid eye movement from one point of interest to another
  - scan paths: moving straight to a target with a short fixation at the target is optimal

Evaluation Techniques

➢ physiological measurements

- ❖ emotional response linked to physical changes
- ❖ these may help determine a user's reaction to an interface
- ❖ measurements include:
  - heart activity, including blood pressure, volume and pulse.
  - activity of sweat glands: Galvanic Skin Response (GSR)
  - electrical activity in muscle: electromyogram (EMG)
  - electrical activity in brain: electroencephalogram (EEG)
- ❖ some difficulty in interpreting these physiological responses - more research needed

Evaluation Techniques

➢ Factors distinguishing evaluation techniques

- The stage in the cycle at which the evaluation is carried out
- The style of evaluation
- The level of subjectivity or objectivity of the technique
- The type of measures provided
- The information provided
- The immediacy of the response
- The level of interference implied
- The resources required.

Evaluation Techniques

➢ Choosing an Evaluation Method

when in process:	design vs. implementation
style of evaluation:	laboratory vs. field
how objective:	subjective vs. objective
type of measures:	qualitative vs. quantitative
level of information:	high level vs. low level
level of interference:	obtrusive vs. unobtrusive
resources available:	time, subjects, equipment, expertise

Evaluation Techniques

➢ Classification of analytic evaluation techniques

	Cognitive walkthrough	Heuristic evaluation	Review based	Model based
Stage	Throughout	Throughout	Design	Design
Style	Laboratory	Laboratory	Laboratory	Laboratory
Objective?	No	No	As source	No
Measure	Qualitative	Qualitative	As source	Qualitative
Information	Low level	High level	As source	Low level
Immediacy	N/A	N/A	As source	N/A
Intrusive?	No	No	No	No
Time	Medium	Low	Low–medium	Medium
Equipment	Low	Low	Low	Low
Expertise	High	Medium	Low	High

Evaluation Techniques

- **Classification of experimental and query evaluation techniques**

	Experiment	Interviews	Questionnaire
Stage	Throughout	Throughout	Throughout
Style	Laboratory	Lab/field	Lab/field
Objective?	Yes	No	No
Measure	Quantitative	Qualitative/ quantitative	Qualitative/ quantitative
Information	Low/high level	High level	High level
Immediacy	Yes	No	No
Intrusive?	Yes	No	No
Time	High	Low	Low
Equipment	Medium	Low	Low
Expertise	Medium	Low	Low

Universal Design

- **Universal Design :-** is about designing systems so that they can be used by anyone in any circumstance.

Multi-Sensory Systems

- More than one sensory channel in interaction
  - e.g. sounds, text, hypertext, animation, video, gestures, vision
- Used in a range of applications:
  - particularly good for users with special needs, and virtual reality

Universal Design

- Will cover
  - general terminology
  - speech
  - non-speech sounds
  - handwriting
- considering applications as well as principles

Universal Design

➢ Universal Design Principles

- ❖ Equitable use
- ❖ Flexibility in use
- ❖ Simple and intuitive to use
- ❖ Perceptible information
- ❖ Tolerance for error
- ❖ Low physical effort
- ❖ Size and space for approach and use

Universal Design

➢ Usable Senses

- The 5 senses (sight, sound, touch, taste and smell) are used by us every day
  - each is important on its own
  - together, they provide a fuller interaction with the natural world
- Computers rarely offer such a rich interaction
- Can we use all the available senses?
  - ideally, yes
  - practically – no
- We can use • sight • sound • touch (sometimes)
- We cannot (yet) use • taste • smell

Universal Design

➢ Multi-modal vs. Multi-media

- Multi-modal systems
  - use more than one sense (or mode) of interaction

    e.g. visual and aural senses: a text processor may speak the words as well as echoing them to the screen
- Multi-media systems
  - use a number of different media to communicate information

    e.g. a computer-based teaching system: may use video, animation, text and still images: different media all using the visual mode of interaction; may also use sounds, both speech and non-speech: two more media, now using a different mode

Universal Design

➢ Speech

- ❖ Human beings have a great and natural mastery of speech
  - makes it difficult to appreciate the complexities but
  - it's an easy medium for communication

Universal Design

➢ Structure of Speech

- ❖ Phonemes
  - 40 of them
  - basic atomic units
  - sound slightly different depending on the context they are in, these larger units are …
- ❖ Allophones
  - all the sounds in the language
  - between 120 and 130 of them
  - these are formed into …
- ❖ Morphemes
  - smallest unit of language that has meaning.

Speech (cont'd)

- Other terminology:
  - Prosody
    - alteration in tone and quality
    - variations in emphasis, stress, pauses and pitch
    - impart more meaning to sentences.
  - Co-articulation
    - the effect of context on the sound
    - transforms the phonemes into allophones
  - Syntax – structure of sentences
  - Semantics – meaning of sentences

Speech Recognition Problems

- Different people speak differently:
  - accent, intonation, stress, idiom, volume, etc.
- The syntax of semantically similar sentences may vary.
- Background noises can interfere.
- People often “ummm.....” and “errr.....”
- Words not enough - semantics needed as well
  - requires intelligence to understand a sentence
  - context of the utterance often has to be known
  - also information about the subject and speaker

e.g. even if “Errr.... I, um, don’t like this” is recognised, it is a fairly useless piece of information on it’s own

➢ The Phonetic Typewriter

- Developed for Finnish (a phonetic language, written as it is said)
- Trained on one speaker, will generalise to others.
- A neural network is trained to cluster together similar sounds, which are then labelled with the corresponding character.
- When recognising speech, the sounds uttered are allocated to the closest corresponding output, and the character for that output is printed.
  - requires large dictionary of minor variations to correct general mechanism
  - noticeably poorer performance on speakers it has not been trained on

Universal Design

➢ The Phonetic Typewriter (ctd)

a a a ah h æ æ ø ø e e e

o a a h r æ l ø y y j i

o o a h r r r g g y j i

o o m a r m n m n j i i

l o u h v vm n n h hj j j

l u v v p d d t r h hi j

. . u v tk k p p p r k s

. . v k pt t p t p h s s

Universal Design

➢ Speech Recognition: useful?

- Single user or limited vocabulary systems

  e.g. computer dictation

- Open use, limited vocabulary systems can work satisfactorily

  e.g. some voice activated telephone systems

- general user, wide vocabulary systems …

  … still a problem

- Great potential, however
  - when users hands are already occupied

    e.g. driving, manufacturing

  - for users with physical disabilities
  - lightweight, mobile devices

Universal Design

➢ Speech Synthesis

- The generation of speech
- Useful
  - natural and familiar way of receiving information
- Problems
  - similar to recognition: prosody particularly
- Additional problems
  - intrusive - needs headphones, or creates noise in the workplace
  - transient - harder to review and browse

Universal Design

➢ Speech Synthesis: useful?

- Successful in certain constrained applications when the user:
  - is particularly motivated to overcome problems
  - has few alternatives
- Examples:
- screen readers
  - read the textual display to the user
    utilised by visually impaired people
- warning signals
  - spoken information sometimes presented to pilots whose visual and haptic skills are already fully occupied

Universal Design

➢ Non-Speech Sounds

- commonly used for warnings and alarms
- Evidence to show they are useful
  - fewer typing mistakes with key clicks
  - video games harder without sound
- Language/culture independent, unlike speech

➢ Non-Speech Sounds: useful?

- Dual mode displays:
  - information presented along two different sensory channels
  - redundant presentation of information
  - resolution of ambiguity in one mode through information in another
- Sound good for
  - transient information
  - background status information

e.g. Sound can be used as a redundant mode in the Apple Macintosh; almost any user action (file selection, window active, disk insert, search error, copy complete, etc.) can have a different sound associated with it.

Universal Design

➢ Auditory Icons

- Use natural sounds to represent different types of object or action
- Natural sounds have associated semantics which can be mapped onto similar meanings in the interaction

  e.g. throwing something away

  ~ the sound of smashing glass
- Problem: not all things have associated meanings
- Additional information can also be presented:
  - muffled sounds if object is obscured or action is in the background
  - use of stereo allows positional information to be added

Universal Design

➢ Sonic Finder for the Macintosh

- items and actions on the desktop have associated sounds
- folders have a papery noise
- moving files – dragging sound
- copying – a problem …

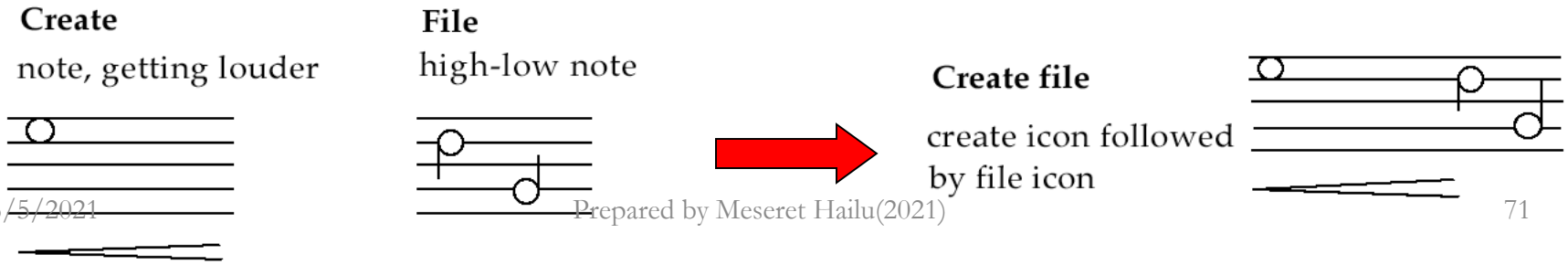
  sound of a liquid being poured into a receptacle

  rising pitch indicates the progress of the copy
- big files have louder sound than smaller ones

Universal Design

➢ Earcons

- Synthetic sounds used to convey information
- Structured combinations of notes (motives) represent actions and objects
- Motives combined to provide rich information
  - compound earcons
  - multiple motives combined to make one more complicated earcon



➢ Earcons (ctd)

- family earcons
  - similar types of earcons represent similar classes of action or similar objects: the family of "errors" would contain syntax and operating system errors
- Earcons easily grouped and refined due to compositional and hierarchical nature
- Harder to associate with the interface task since there is no natural mapping

Universal Design

➢ Touch

- haptic interaction
  - cutaneous perception
    - tactile sensation; vibrations on the skin
  - kinesthetics
    - movement and position; force feedback
- information on shape, texture, resistance, temperature, comparative spatial factors
- example technologies
  - electronic braille displays
  - force feedback devices e.g. Phantom
    - resistance, texture

➢ Handwriting recognition

- Handwriting is another communication mechanism which we are used to in day-to-day life
- Technology
  - Handwriting consists of complex strokes and spaces
  - Captured by digitising tablet
    - strokes transformed to sequence of dots
  - large tablets available
    - suitable for digitising maps and technical drawings
  - smaller devices, some incorporating thin screens to display the information
    - PDAs such as Palm Pilot
    - tablet PCs

Universal Design

➢ Handwriting recognition (ctd)

- Problems
  - personal differences in letter formation
  - co-articulation effects
- Breakthroughs:
  - stroke not just bitmap
  - special ‘alphabet’ – Graffeti on PalmOS
- Current state:
  - usable – even without training
  - but many prefer keyboards!

Universal Design

➢ Gesture

- applications
  - gestural input - e.g. “put that there”
  - sign language
- technology
  - data glove
  - position sensing devices e.g MIT Media Room
- benefits
  - natural form of interaction - pointing
  - enhance communication between signing and non-signing users
- problems
  - user dependent, variable and issues of co-articulation

Universal Design

➢ Users with disabilities

- visual impairment
  - screen readers, SonicFinder
- hearing impairment
  - text communication, gesture, captions
- physical impairment
  - speech I/O, eyegaze, gesture, predictive systems (e.g. Reactive keyboard)
- speech impairment
  - speech synthesis, text communication
- dyslexia
  - speech input, output
- autism
  - communication, education

Universal Design

➢ … plus …

- age groups
  - older people e.g. disability aids, memory aids, communication tools to prevent social isolation
  - children e.g. appropriate input/output devices, involvement in design process
- cultural differences
  - influence of nationality, generation, gender, race, sexuality, class, religion, political persuasion etc. on interpretation of interface features
  - e.g. interpretation and acceptability of language, cultural symbols, gesture and colour

- End of unit Seven

Evaluation Techniques and Universal Design