# First Design & Evaluation

## Unit 14c

Zeno Menestrina, MSc
*zeno.menestrina@unitn.it*
Prof. Antonella De Angeli, PhD
*antonella.deangeli@unitn.it*

# Group work - calendar

- Week V – game ideation
- Weeks VI, VII – PACT analysis and User requirements
- Week VIII – first design
- Week IX – prototyping
- Week X – presentation
- Week XI, XII – evaluation (from heuristics and users)
- Weeks XIII, XIV – re-design

# Presentation (1)

- Around 13 groups

- Starting from Tuesday 22th

  (if we need more time, we continue on Thursday 24th)

# Presentation (2)

- ~5 minutes presentation
  1. Game idea
  2. First data collection
     (how you made it, which info you gathered)
  3. First design
  4. Future plans
     (how you'll  evaluate it)

- Feedbacks by Prof. De Angeli, Andrea Conci and I (Zeno)

# Post-presentation

- Upload a short description and prototypes to a platform for idea generation and discussion
- Provide and receive feedbacks from your colleagues
- Considered (positively) as part of the coursework
- More details during next lectures

# Next lectures

- A couple of theoretical lectures
- Workshops on the coursework

# Evaluation

- What is the purpose of your evaluation?
  - Specific features?
  - Overall design?
  - User experience?
  - Specific usability issues?

# Data gathering

- Documentation
- Similar products
- Observation
- **Interviews**
- **Questionnaires**

# Interviews

- Structured, unstructured or semi-structured
- Good for exploring issues
- Time consuming -> may be infeasible to talk to everyone

# Focus Group

- Focus group
  - Group interviews
  - Good at gaining a consensus view and/or highlighting areas of conflict

# Interview - how

- What do you want to know?
  - Plan a set of central questions
  - Avoid leading questions to not bias
  - Focus the interview

- Let user's responses lead follow-up questions
  - Follow interesting leads
  - Vary questions to suit the context

# Interview – tricks (1)

- Prompts
  - Remain silent
  - Repeat the last question
  - Repeat the last few words by the interviewee
- Probes
  - Verbal
    - ‘could you give me some examples of that’
    - ‘would an example of that be .....’
    - Could you give me a bit more details on ....
  - Design
    - Interfaces
    - Scenario, storyboard

# Interview – tricks (1)

- Checks
  - ‘If I can summarise what I think you’ve said...’
  - ‘What this means, then is that,...’
  - So let me check if I have understood you correctly’

# Interview – analysis

- Look for key events/patterns of behavior that drive the activity
- Recurring patterns or themes
  - Emergent from data
  - Emergent from theory
- Categorizing data
  - Categorization scheme may be emergent or pre-specified
- Look for critical incidents
  - Helps to focus in on key events

# Questionnaires

- Predefined set of questions in a predetermined order
- Provide data that can be analysed to identify patterns and relationships within the answers
- Qualitative vs. quantitative data

# Characteristics

- Reliability (reproducibility factor):
  - test-retest reliability: stability over time
  - internal reliability: stability over items
- Effective range in the scale (sensibility):
  - scale must be chosen according to variable peculiarities.
- Validity:
  - measures must reflect what it is investigated.

# Sampling

- Random Sampling
  - People selected at random from a population: All NHS patients, NHS Patient within a Primary Care Trust
- Systematic Sampling
  - People/events at regular intervals from random sample: NHS hypertensive patients remotely tele-monitored at regular intervals when presenting themselves at hospital
- Stratified Sampling
  - Choosing People/events that map stratifications in the whole
  - population

    Matching patients from various socioeconomic classes with diabetes remotely monitored on line, via phone.
- Snow-ball sampling
  - Self-selected sample: Patients with hypertension forwarding the survey to acquaintances with similar characteristics

# Questionnaire design (1)

- Questionnaires must be properly designed to elicit the answers you want
- Answers should be valid and meet the researchers' need
- Each question has to effectively elicit an answer that contributes to addressing the overall research question

# Questionnaire design (2)

- How long does it take to complete?
- Arethe instructions clear?
- Are any questions ambiguous?
- Areany questions objectionable?
- Is the layout clear and easy to follow?
- Are any topics omitted?

# Styles of Questions

- Open-ended questions
  - asks for unprompted opinions
  - good for general subjective information
    - but difficult to analyze rigorously

"Can you suggest any improvements to the interface?"

# Closed questions

- Restrict responses by supplying alternative
- answers  easy to analyze
- watch out for hard to interpret responses!

Do you use computers at work:
O often  O sometimes  O rarely
*vs*
In your typical work day, do you use computers:
O over 4 hrs a day
O between 2 and 4 hrs daily  O between 1and 2 hrs daily
O less than 1 hr a day

# Multi-choice

- Respondents offered a choice of explicit responses

How do you most often get help with the system? (tick one)

✓ O on-line manual O paper manual O ask a colleague

Which types of software have you used? (tick all that apply)

O word processor O data base

✓ O spreadsheet ✓ O compiler

# Ranked

- respondent places an ordering on items in a list
- useful to indicate preferences
- forced choice
- Limit the number of items

Rank the usefulness of these methods of issuing a command
(1 most useful, 2 next most useful..., 0 if not used

__2___command line

__1___menu selection

__3___control key accelerator

# Likert Scales

- User judge a specific statement on a numeric scale
- usually corresponds with agreement or disagreement with a statement

The characters on the computer screen are hard to read

| 1 | | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Strongly agree | agree | neutral | disagree | strongly | disagree |

# Combination

- Combining open-ended and closed questions
  - get specific response, but allows room for user's opinion

It is easy to recover from mistakes:

disagree agree

1 2 3 4 5

comment: *the undo facility is really helpful*

# Semantic differential scale

- Bi-polar attitudes about a concept
- pair of adjectives

The look and feel of the web-site is

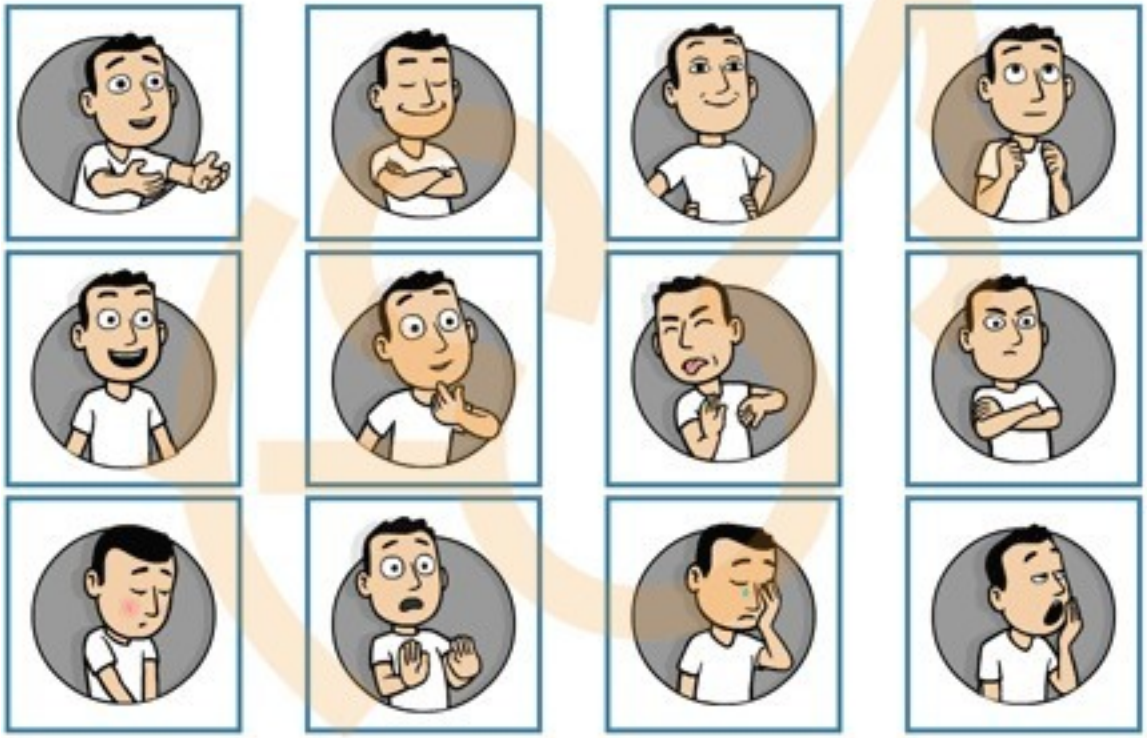
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| exciting | 1 | 2 | 3 | 4 | 5 | 6 | 7 | boring |
| annoying | 1 | 2 | 3 | 4 | 5 | 6 | 7 | pleasing |

# PrEmo

- Emotional responses elicited are difficult to measure because
  - –their nature is subtle (low intensity)
  - –they are often mixed (more than one emotional response at the same time)
- Instead of words, use animated cartoon characters
- Evaluation does not become a rational process
-

# PrEmo

*"To which extent do the feelings expressed by the characters correspond with your own feelings towards the stimulus?"*




http://www.premo-online.com/en/how-does-it-work/

# Data coding

- Direct measurements
  - Data are already in numeric form
- Indirect measurement
  - Need to be coded into a number
    - Code each predefined answer of a
    - questionnaire  Code each time which something
    - happens
      open questions
  - Codes need to be
    - mutually exclusive
    - Exhaustive
    - Consistently applied
  - Code book

# Visualization

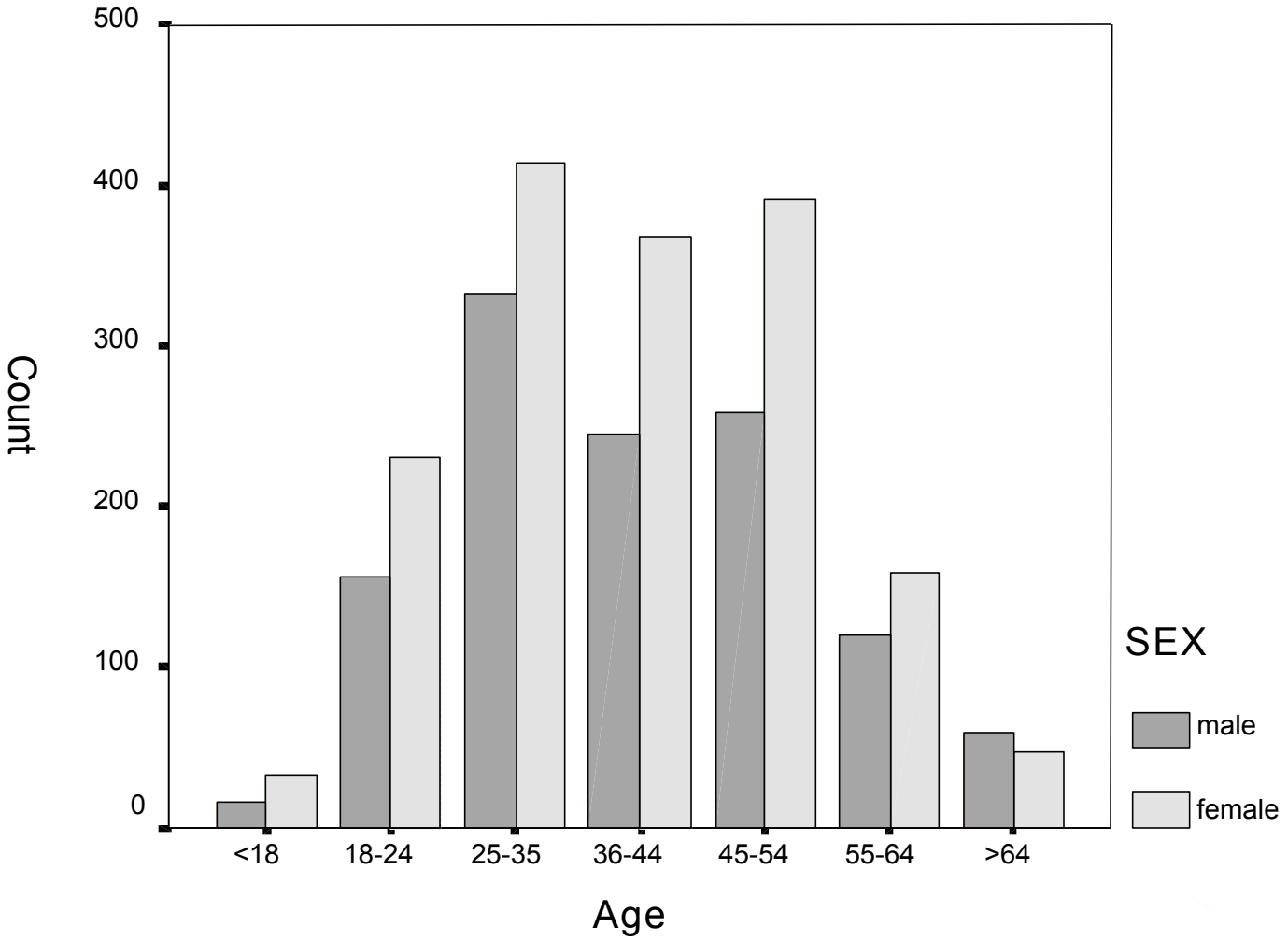
- Tables
  - General summaries
- Bar charts
  - Frequencies - means
- Pie charts
  - Good for showing proportion
- Scatter graph
  - Relationship
- Line graph
  - trends

# Tables

| | Menu-based | | | Metaphor-based | | |
|---|---|---|---|---|---|---|
| | Freq | % | Severity | Freq | % | Severity |
| Poor menu/navigation | 21 | 45 | 2.95 | 47 | 47 | 3.87 |
| Poor graphical design | 13 | 27 | 2.93 | 38 | 38 | 3.97 |
| Poor information | 7 | 15 | 2.50 | 9 | 9 | 3.67 |
| Other | 6 | 13 | 3.33 | 6 | 6 | 3.67 |
| **Total** | **47** | **100** | **2.94** | **100** | **100** | **3.87** |

**Table 2. Statistics of usability problems classified by cause in the two experimental conditions.**

# Bar Charts


500
400
300
200
100
0
Count
<18
18-24
25-35
36-44
45-54
55-64
>64
Age
SEX
male
female

# Line graphs

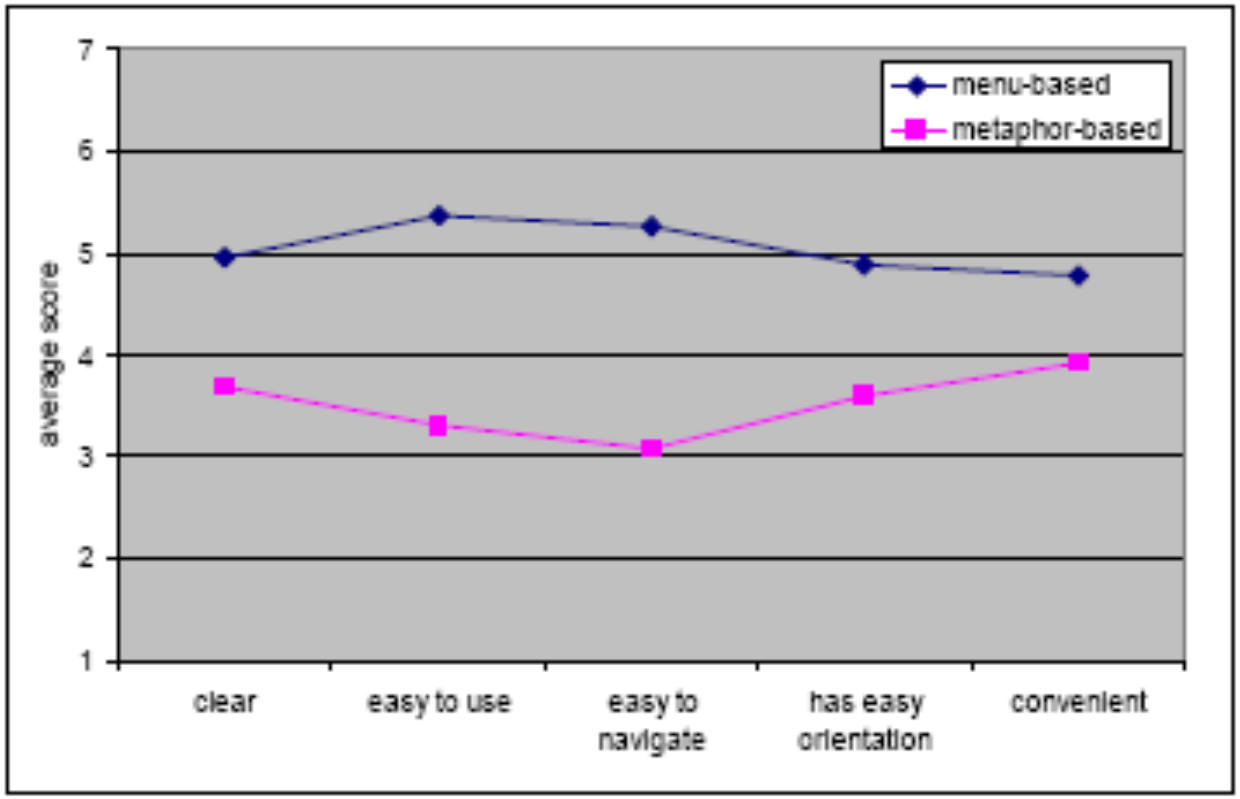



**Figure 3: Usability ratings as a function of experimental condition**

# Pie Charts

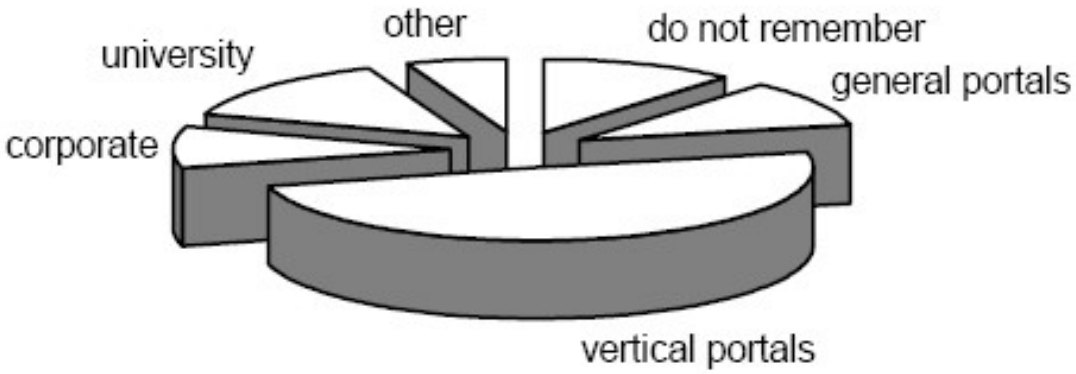



Figure 4. Type of site used to seek advice

# Visualization

- Be easy to read
- Have a title / labels/
- Be referenced in the paper
- Give information about the units represented in it
- Displays enough information but not too much

# Presenting the findings

- Only make claims that your data can support
- The best way to present your findings depends on
  - the audience,
  - the purpose,
  - the data gathering and analysis undertaken
- Graphical representations may be appropriate for presentation

# Observation in usability testing

- More objective
- In usability lab equipment is in place
- Recording is continuous
- Analysis & observation almost simultaneous
- Care needed to avoid drowning in data
- Analysis can be coarse or fine grained
- Video clips can be powerful for reporting data

# Simple observation

- User is given a task
- Evaluator just watches the user

- Problem

– does not give insight into the user's decision process or attitude

# Think aloud method

- Users speak their thoughts while doing the task
  - what they are trying to do
  - why they took an action
  - how they interpret what the system did
  - gives insight into the user's mental model of the system, but
    - may alter the way users do the task
    - unnatural (awkward and uncomfortable)
    - hard to talk if they are concentrating

# Constructive interaction method

- Two people work together on a task
  - monitor their normal conversation
  - removes awkwardness of think-aloud
  - Introduce group variability
- Co-discovery learning
  - use semi-knowledgeable “coach” and novice
  - only novice uses the interface
    - novice ask questions
    - coach responds
  - gives insights into two user groups

# Internet based observations

- On-line communities, social-networking platforms, web 2.0
- The observer joins in the community and participate to it
- The observer only look at the community from the external
- Useful to understand how technology is used in real life
- http://www.wikimedia.org/

# Diaries

- Critical incident diaries
  - write when something goes wrong
- Reflection diaries
  - post-usage
- Sampling activity
  - Timed
- Requires incentives
- Better if supported by different medias (e.g., cameras)

# Decide

- *D*etermine the evaluation *goals*
- *E*xplore specific *evaluation questions*
- *C*hoose the *evaluation paradigm* and *techniques* to answer the questions.
- *I*dentify the *practical issues.*
- *D*ecide how to deal with the *ethical issues*.
- *E*valuate, interpret and present the *data.*

# Which techniques?

- Depends on:
- Amount of time, level of detail and risk associated with the findings
- Knowledge of the analyst
- Kind of task to be studied:
  - Sequential steps or overlapping series of subtasks
  - High or low, complex or simple information?
  - Task for a layman or a skilled practitioner?

# Today

- What is the purpose of your evaluation?
- What do you need to evaluate? Why?

Answer these questions and start defining your evaluation