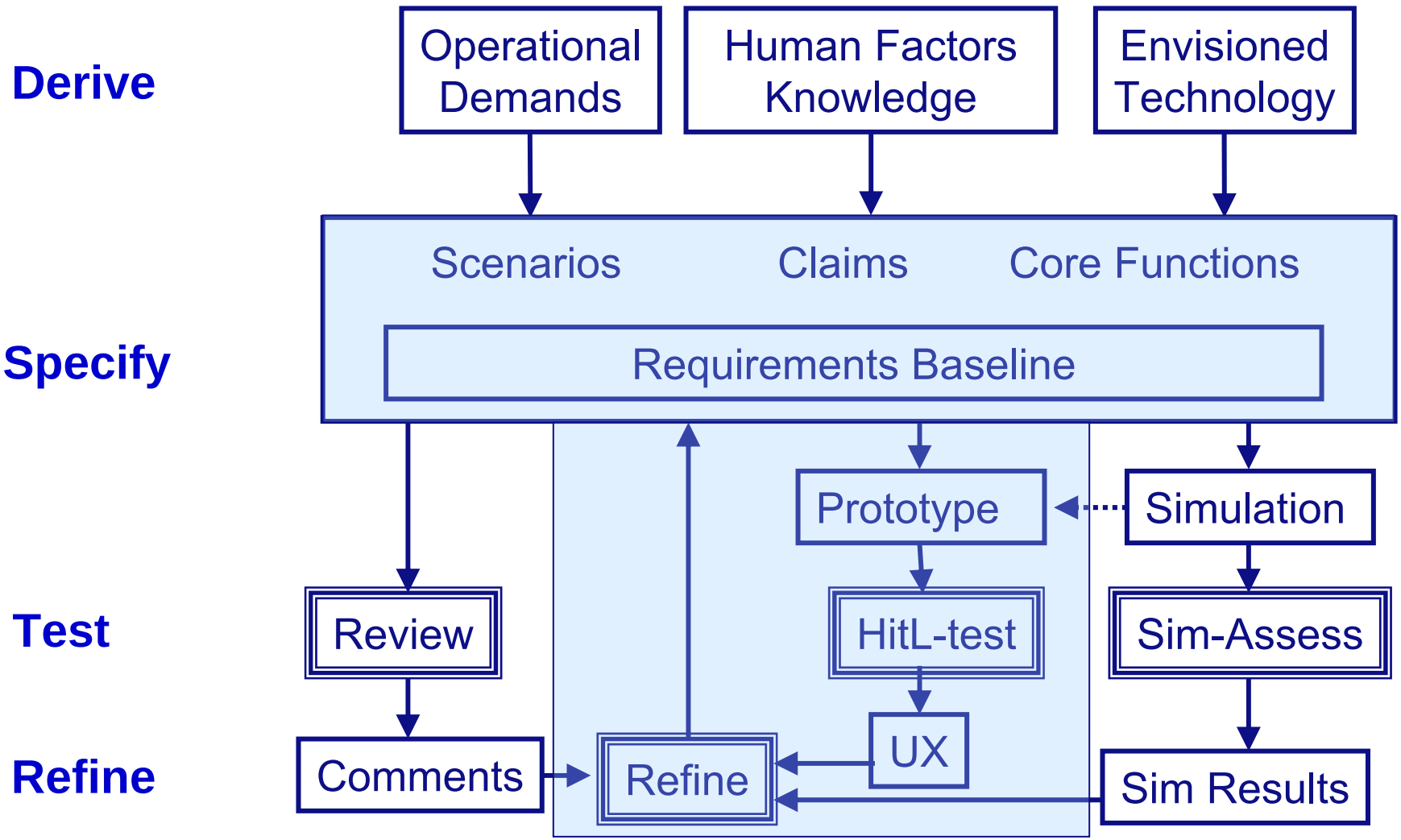
TNO Defensie en Veiligheid

Evaluation in virtual environments

Nanja Smets

TNO | TU Delft

Situated Cognitive Engineering



Outline

- Human in the loop
- Virtual environment

- 4 case studies from 2 different projects
  - 2 with virtual environment on screen
  - Screen vs. 2D Cave
  - 2D Cave

Evaluation methods

- Focus Groups
- User Walkthroughs
- Wizard-of-Oz
- Real-life testing
- Game-based evaluation in a virtual environment

Field testing

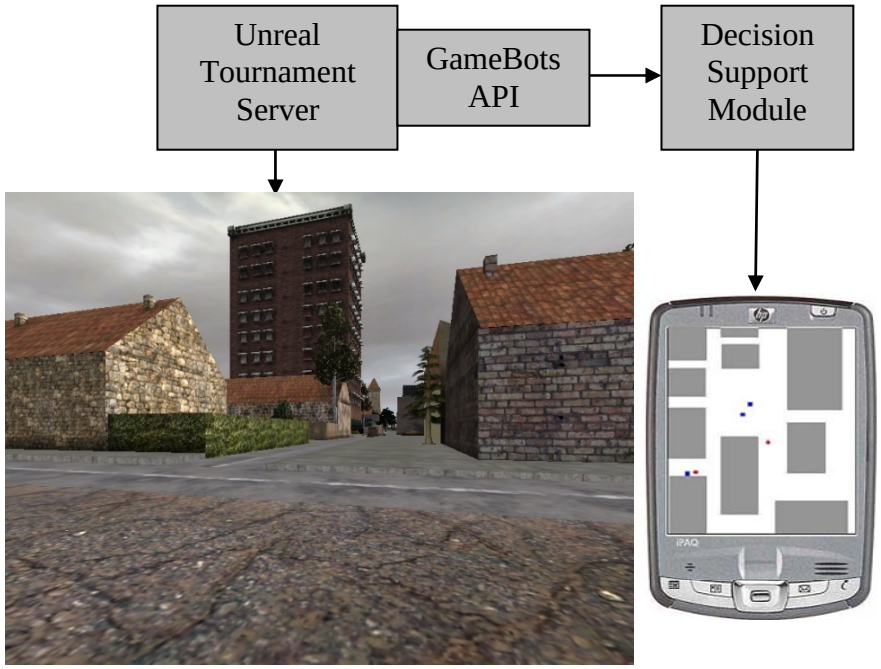
- Realistic
- Expensive
- Complex
- Sometimes risky
- Not always possible

Virtual Environment

- Context effects
- Controllable
- Easy logging
- Inexpensive
- Validation?

- Virtual environments suited for scrutinizing new concepts before bringing them to the field

Virtual environment



Case Study 1: Mobile support for first responders

- First on scene of a crisis

- Use of maps
- Navigation
- Situation Awareness (SA)

Research questions

- Is there a relation between map size and the quality of the situational awareness that is developed by the first responders?
  - A better rescue performance and situation awareness is expected with a “tablet” map size than with a “PDA” map size.
- Does good spatial ability have a substantial positive effect on the number of rescued victims, and is there an interaction effect with map size?
- Is the STE a good environment to evaluate a mobile support tool?
  - What are the advantages and disadvantages of using a STE?

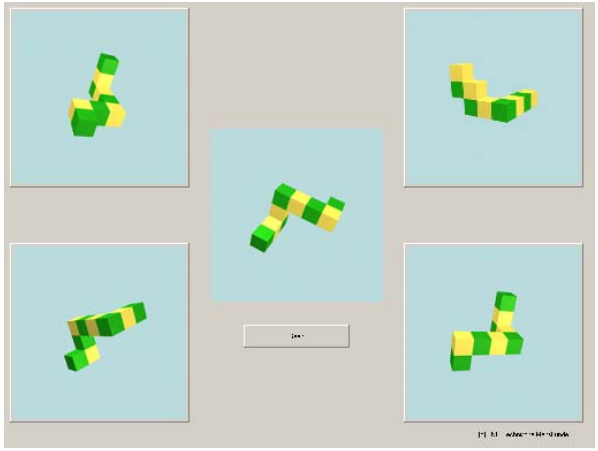
Virtual environment

Materials

- Left screen: support tool
- Right screen: synthetic environment
- Game controller to walk through the world

Design and procedure

- 20 participants
- Between participant design
  - Map: tablet size or PDA size
- Pre-test and post-test in a similar environment to assess practice effect navigation skills
- Spatial test



Task

- Search and rescue victims
- Develop situation awareness (SA): remember important landmarks and events
- Map showing own location and location of victims

Tablet size

PDA size

Dependent variables

- Performance data
  - Number of victims rescued
  - Distance walked in environment
- Subjective data gathered in questionnaires
  - Navigation (PDA, environment, both)
- Situation awareness
  - Number of victims rescued
  - Own location
  - Location of landmarks

SAGAT

- SAGAT – Situation Awarenss Global Assessment Technique (Endsley, Bolté and Jones, 2003)
  - Suspend task at random times, but the same for all participants
  - Question
  - Resume task

- Goal directed task analysis
  - Identify goals of operator
  - Decisions that operator must make to achieve these goals
  - Identify interesting questions about task that covers all levels of SA

Goal directed task analysis

Goal directed task analysis	**Goal:** Rescue all victims **Sub goal:** Find victim **Sub goal:** Find exit of building Verify path to next victim/ exit Place of myself Place of the victim/ exit Obstructions in path Check if information you have is last updated information
SA questions sample set	Indicate on the map which areas are obstructed. Indicate on the map where there was fire. Indicate your location at this moment?

Results (1)

- Is there a relation between map size and the quality of the situational awareness that is developed by the first responders?
  - A better rescue performance and situation awareness is expected with a “tablet” map size than with a “PDA” map size.

→ Participants did have a better overview in the tablet condition of the number of victims they had rescued than in the PDA condition

No other effects were found

- Too many victims in the STE, running around was more efficient than route planning

Results (2)

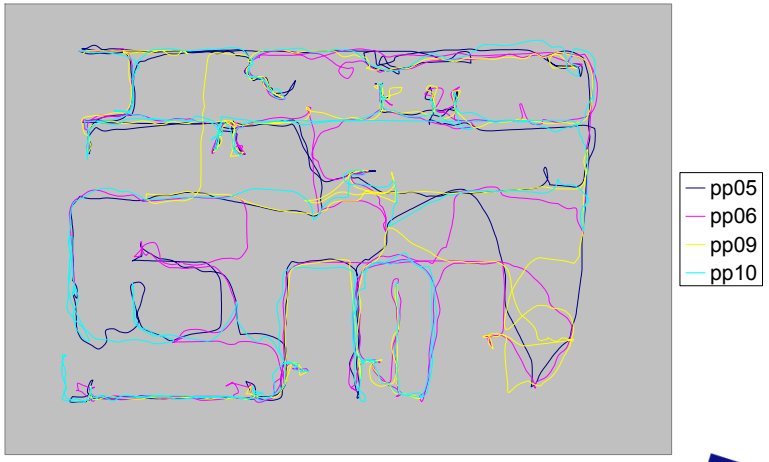
Does good spatial ability have a substantial positive effect on the number of rescued victims, and is there an interaction effect with map size?

→ Results suggested that spatial ability had an effect on the number of rescued victims, which can be attributed to planning opportunities the tablet condition provided.

Results (3)

- Is the STE a good environment to evaluate a mobile support tool?
  → The STE was considered a good environment to evaluate mobile support tools.
    - Easy logging of events
    - Easy to extract data and process it into useful visualizations to interpret the results.
    - Participants liked the experiment and were motivated
    - None of the participants experienced motion sickness

    - Validation issue
      - Is SA measurable in a STE?
      - SA was lower than expected



Case study 2: mobile support for first responders

- Audio and visual modalities are overloaded
- Solution: Mobile tactile display
  - Navigation
  - Situation Awareness

- Navigation by soldiers in forested areas
  - Higher walking speed
  - Reached more targets
  - Decreased visual load
  - More difficult to build SA
    - Lack of information on position and distance to waypoints

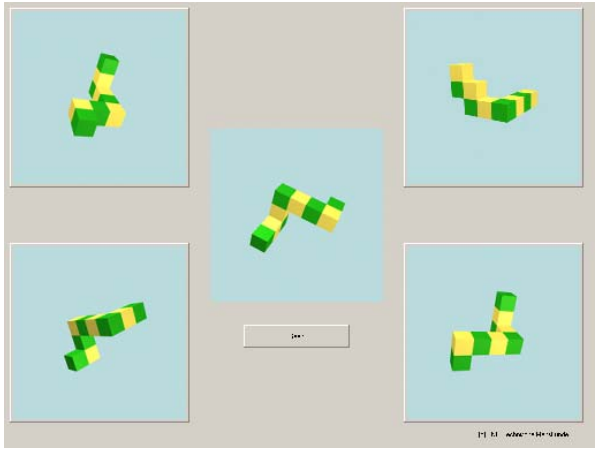
Hypotheses

- Generic task: Someone needs help. Responder has to go to the victim and return to starting point as quickly as possible. Victim's location is presented on map and tactile.

- Heading-up maps are more effective for navigational tasks, e.g. going to a victim.
- North-up maps better support building SA
- Tactile support improves navigation speed
- Tactile support decreases SA

Design and procedure

- 42 participants
- Between participant design
  - Map: heading-up or north-up
  - Tactile feedback: on or off
- Pre-test and post-test in a similar environment to assess practice effect navigation skills
- Spatial test



Task

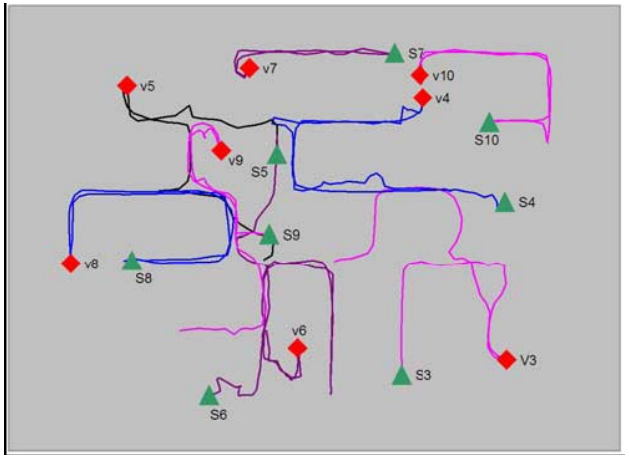
- Go to victim as fast as possible using provided support
- Return to starting point without support to measure SA
- Participant was ‘teleported’ to a new location on the map
- Task started over again, with new starting point

Dependent variables

- Performance data
  - Navigation speed
- Satisfaction
  - Questionnaire
- Situation awareness
  - Subjective questionnaire
  - Performance measures
    - Time to walk from victim to the starting point
    - Number of times starting point was found

Results (1)

- Heading-up maps are more effective for navigational tasks, e.g. going to a victim.
  - Map orientation significantly improved the performance. Victims were rescued faster with a heading-up map than with a north-up map.



Results (2)

- Tactile support improves navigation speed
  - Tactile display in combination with a heading-up map does not improve performance, but when added to a north-up map performance was improved significantly.

  → For tasks where having good SA is important, north-up maps in combination with tactile waypoint information may be a good design option.

Results (3)

- North-up maps better support building SA
  - No significant results were found
- Tactile support decreases SA
  - The data suggest that tactile feedback lowers SA
  - Subjective data showed that participants had the feeling that the tactile suit had an impact on the situation awareness

  → Whether SA is really lowered by adding tactile information requires furthers study.

- Not answered satisfactory: Is responder's SA lowered by adding tactile information?

Case study 3:
Mission Execution Crew Assistant (MECA)

- Internship by Martijn Abbing (University of Utrecht)

- MECA: The objective of MECA is to empower the cognitive capacities of human-machine teams during planetary exploration missions in order to cope autonomously with unexpected, complex and potentially hazardous situations.

- Context of evaluation
- Game-based evaluation

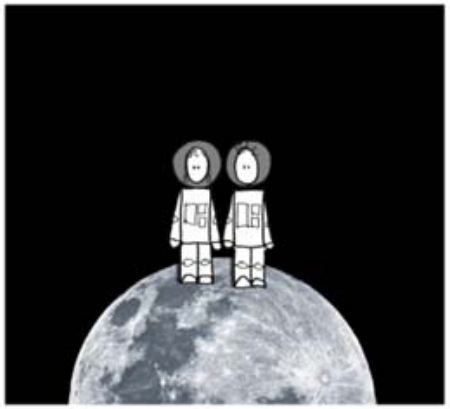
Research questions

- Does the evaluation of prototype systems in rich evaluation conditions contribute to producing a more elaborate, in-depth and realistic user experience (affective and cognitive), better task involvement and more realistic experience of events in the scenario?

- Does the evaluation condition influence the evaluation results of the tested system, i.e. does the context of evaluation have an influence on how the system is used.

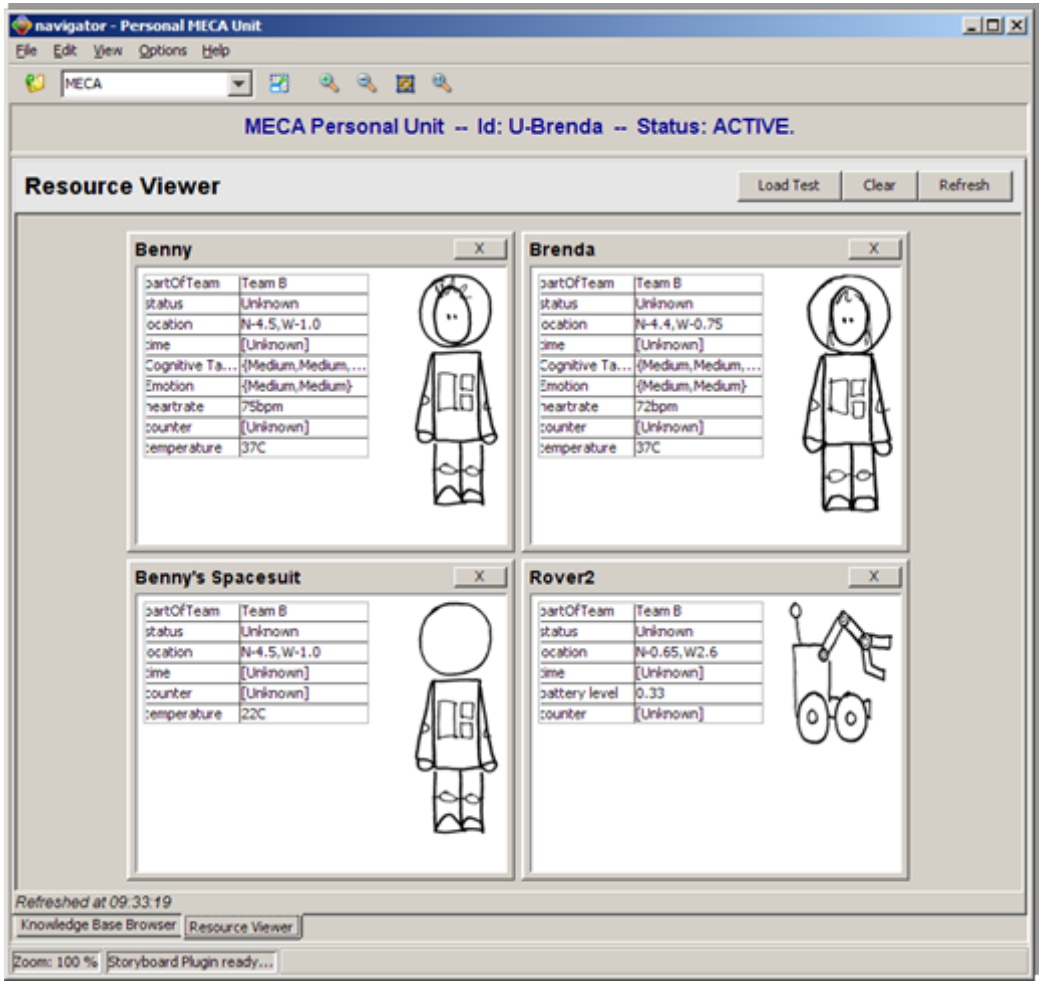
Simulated environment

Without simulated environment

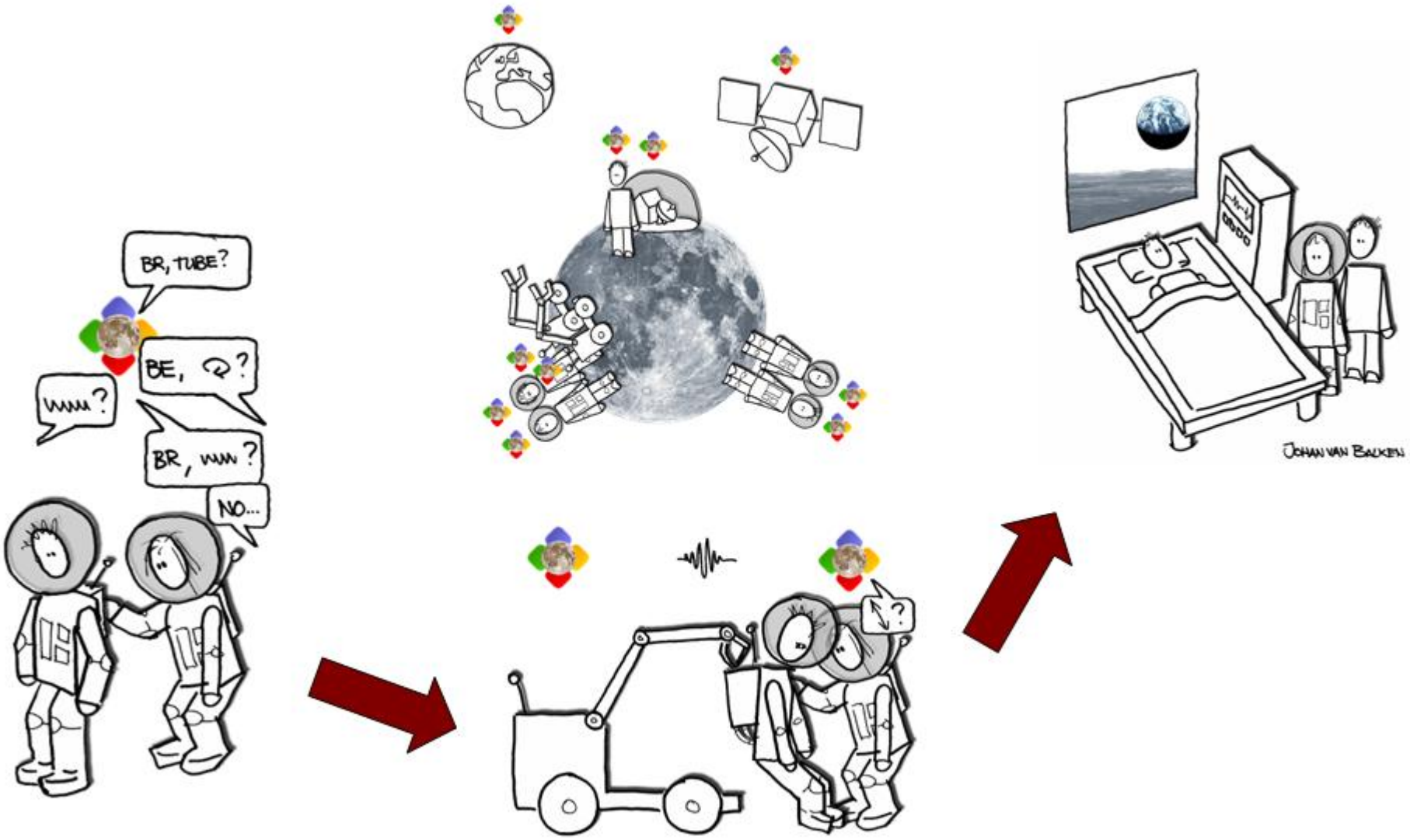


Brenda Ross (you) and Benny Parker are on the surface of the moon, on a EVA (Extra Vehicular Activity). Your goal is to collect rocks for research purposes.

MECA prototype



MECA Scenario

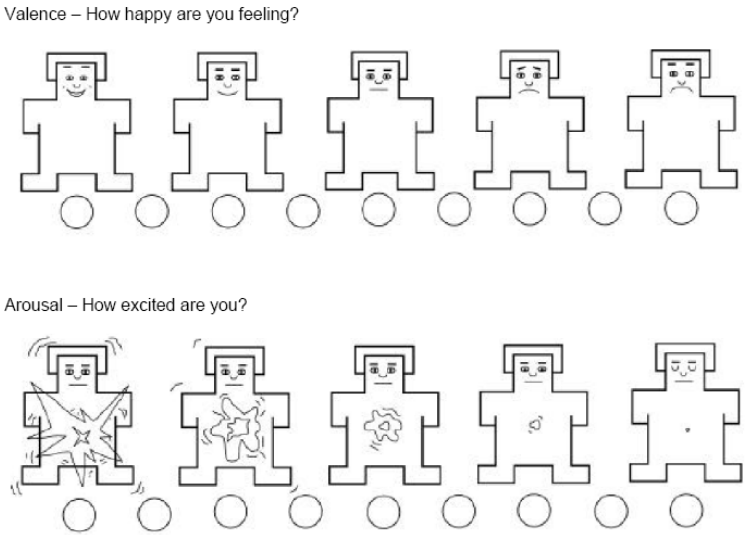


Experiment

- Between subjects design
- 25 participants
- Scenario
- Wizard of Oz
- 7 tasks
  - Blended into scenario
  - Conversational messages
  - Example:
    - Search for a procedure on how to request a rover
    - Monitor temperature value of colleague

Dependent variables (1)

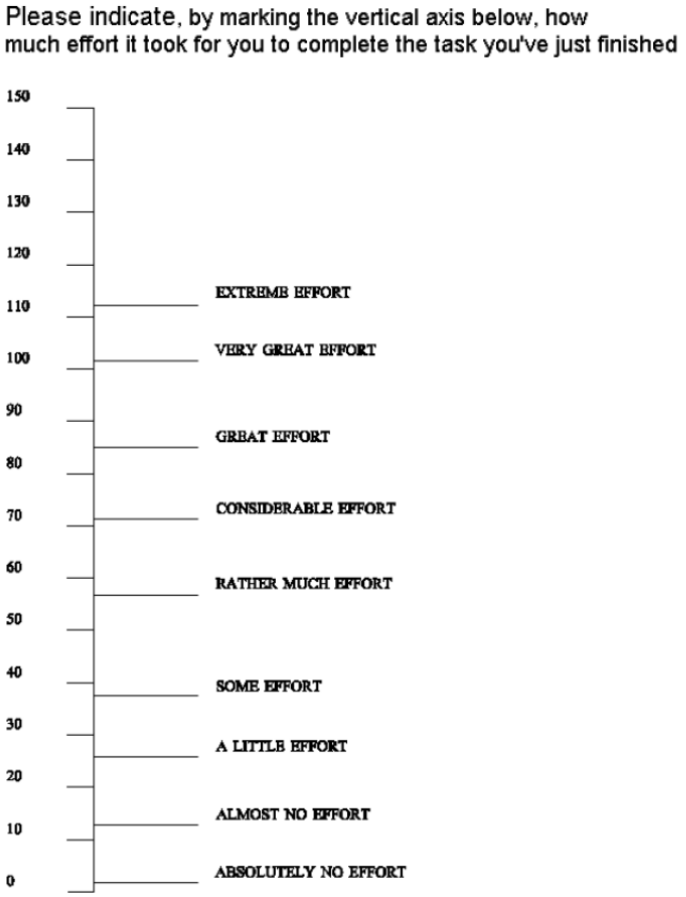
- Emotion (arousal & valence)
  - Self-Assessment Manikin (7 moments during experiment)
- Presence
  - Igroup Presence Questionnaire
- Overall effects (Questionnaires after experiment)
  - Trust
  - Satisfaction
  - Acquired knowledge using the system
  - User feedback on system functionalities



Self Assessment Manakin (Bradley & Lang, 1994)

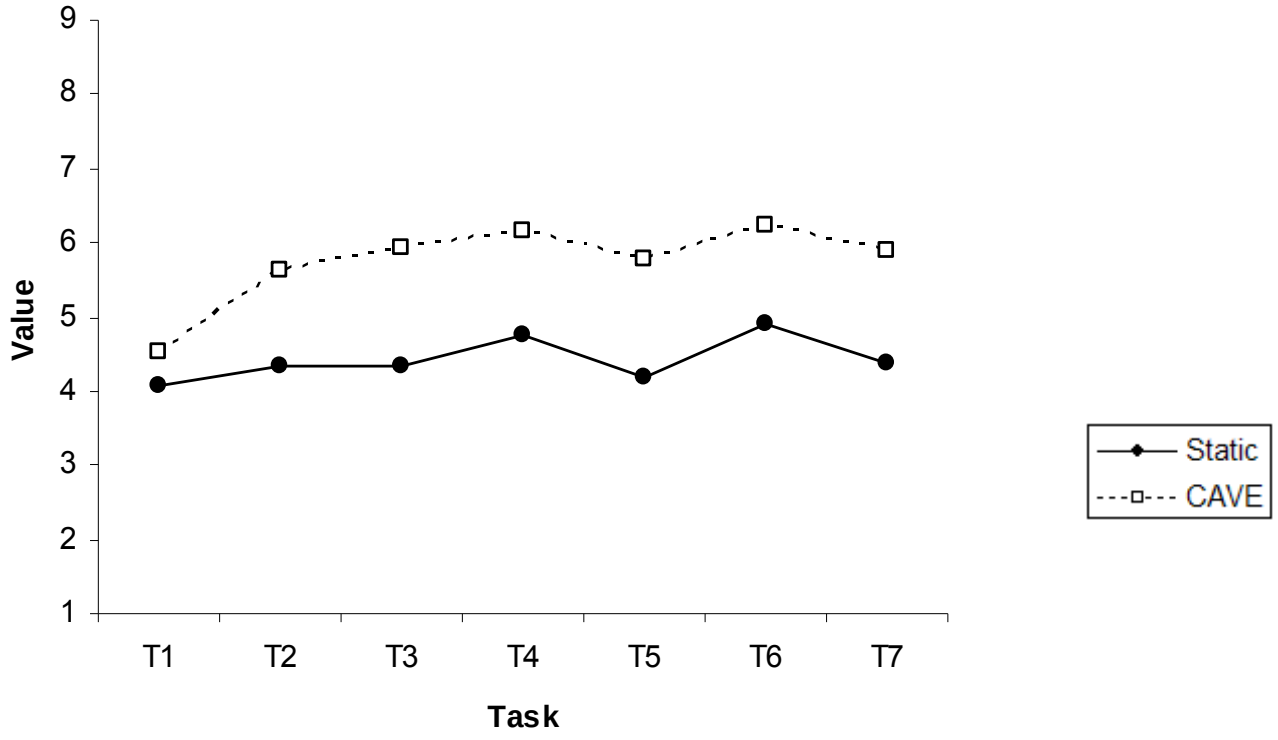
Dependent variables (2)

- Mental effort
  - Rating Scale for Mental Effort (7 moments during experiment, after each task)
- Situation Awareness
  - Scenario and environment related questions (after experiment)
- Performance
  - Task times (failed after 3 minutes)

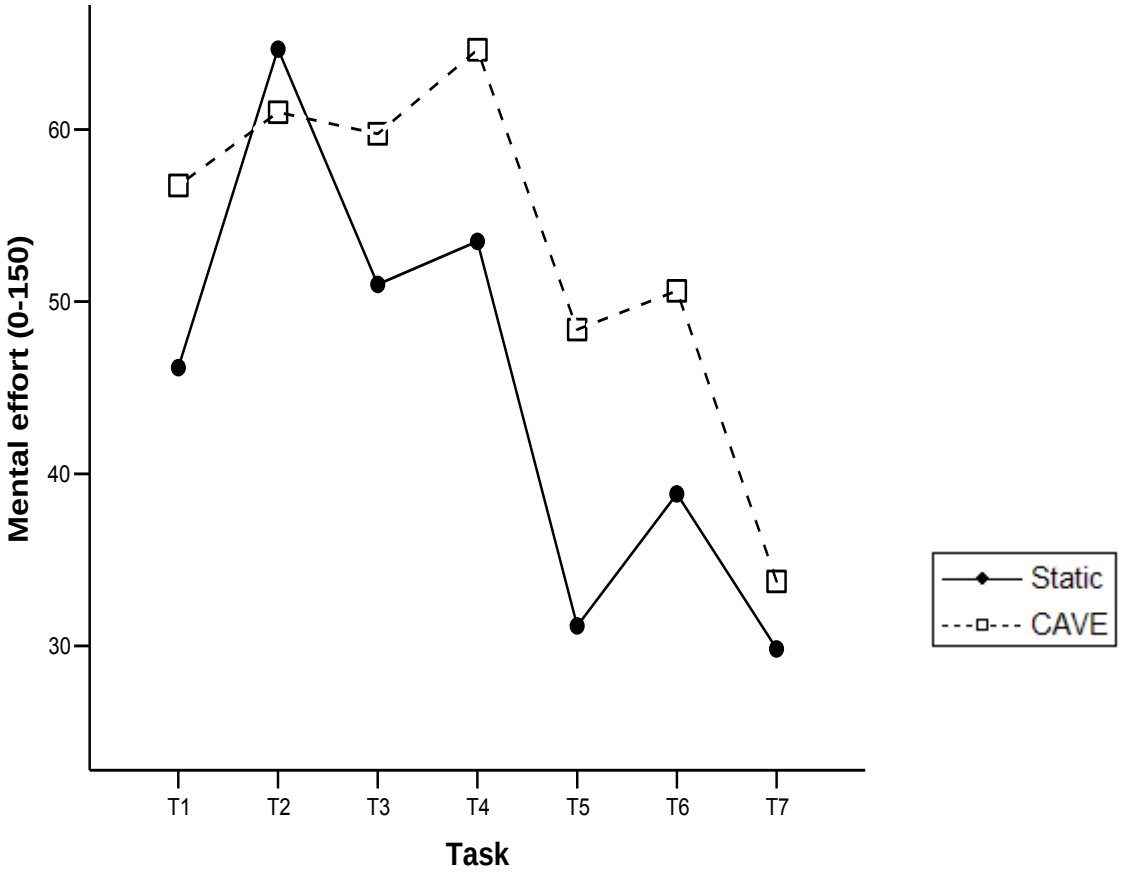


Rating scale for mental effort (Zijlstra, 1993)

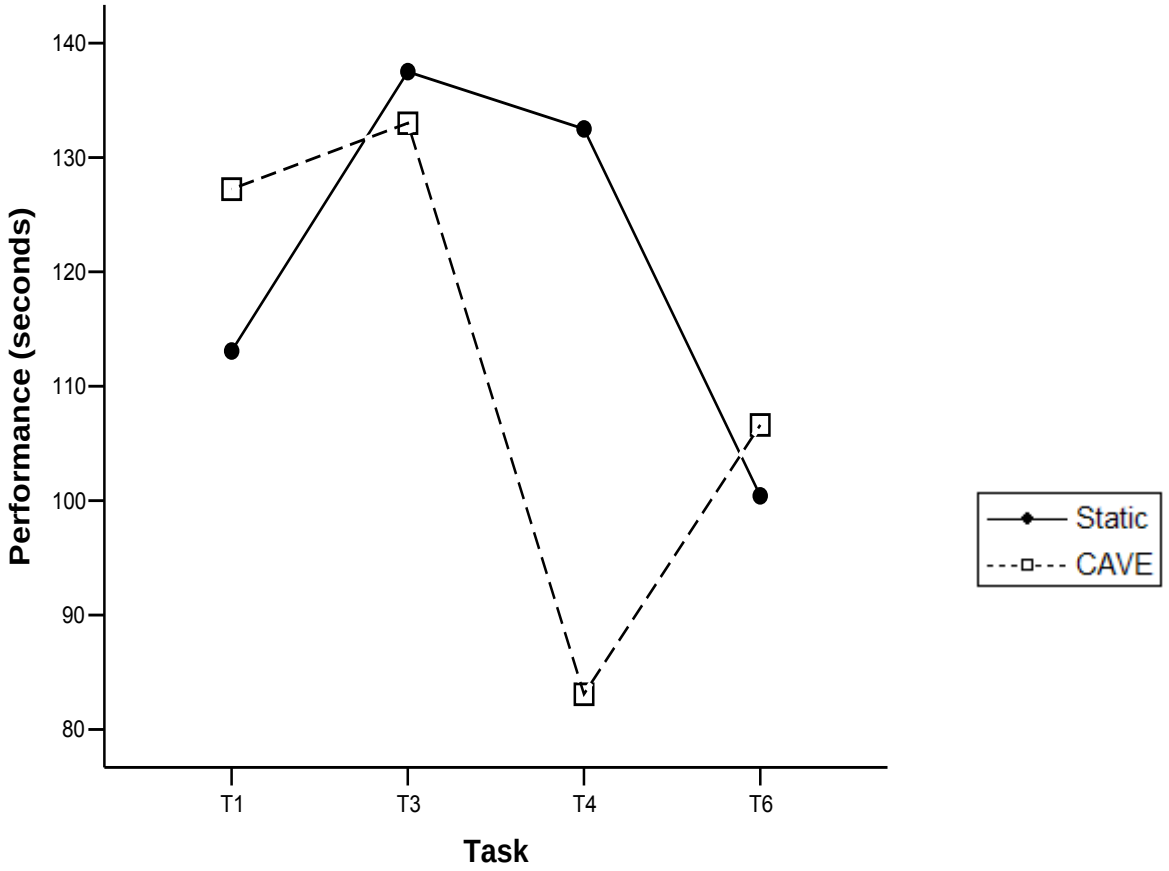
Results (1) arousal



Results (2) mental effort



Results (3) Performance



Results (4)

- Game-based evaluation:
  - More realistic performance
    - Higher arousal levels
    - Higher mental effort
    - Faster performance
  - “Realistic” subjective emotion after critical event
  - More intense feeling of presence
  - Better situation awareness

Conclusions

- Advantages of game-based evaluation
  - Simulate eventual context of use (e.g. space)
  - The results showed that game-based evaluation has additional value to the outcomes of the evaluation
  - Different stages of the development process
- These results are important for developing a support system for people in high demand situations

Case study 4: MECA Evaluation – Simulation based

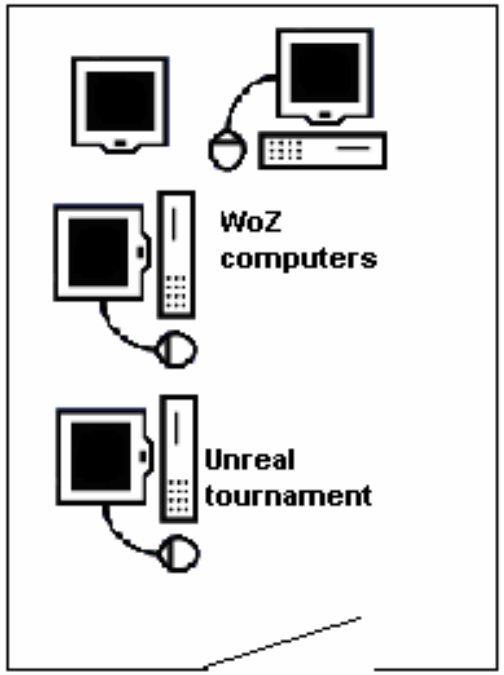
- October 2007, TNO Soesterberg
- 15 participants
  - Including 4 submarine operators

- Objective
  - refine and validate the RB

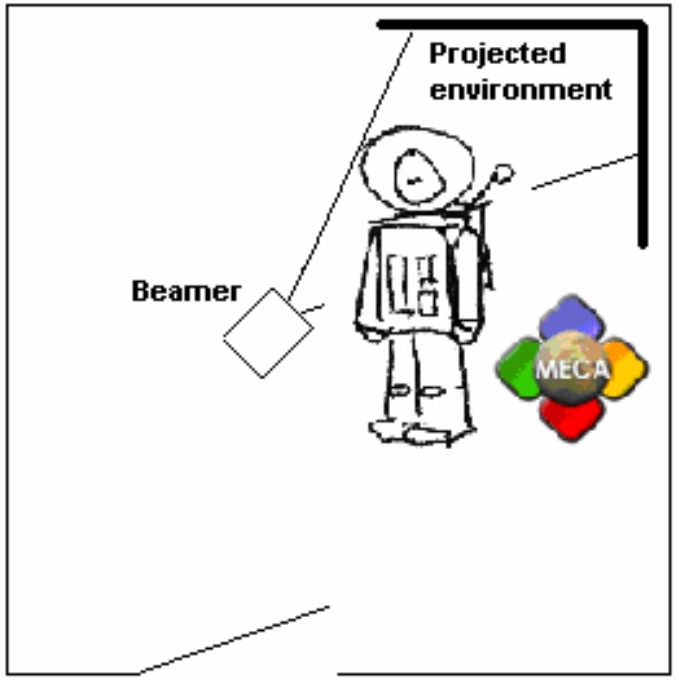
Claims

Scenario	Claims
1. Benny and Brenda are performing rock collecting procedure. Herman is in the habitat performing cardiopres procedure and keeping an overview of all resources.	
2. Benny's spacesuit heater fails.	Emotion ↓ Situation awareness ↑
3. Benny and MECA unit detect and diagnose problem together.	Effectiveness ↑ Efficiency ↑ Easy-to-learn ↑
4. Brenda helps Benny by looking if there is ice on his suit.	Emotion ↓ Situation awareness ↑
5. MECA predicts hypothermia and calls for help from other MECA-equipped entities (astronauts, equipment, habitats, facilities, and rovers).	Situation awareness ↑ Satisfaction ↑
6. MECA advises Team B to start walking towards the habitat and be picked up by a rover on the way. Benny accepts the advice and Team B starts walking.	Trust ↑
7. Herman and MECA select rover to pick up Team B	Situation awareness ↑
8. Herman in the habitat prepares to receive the astronaut.	Easy-to-learn ↑
9. MECA informs astronaut (and other entities) of plan.	Situation awareness ↑
10. Benny faints earlier than predicted.	Trust ↓
11. Brenda, MECA and rover devise way to pick up Benny.	Easy-to-learn ↑ Emotion ↑ Effectiveness ↑ Efficiency ↑
12. Rover transports Benny and Brenda to the habitat.	Situation awareness ↑

Prototyping, simulation and testing

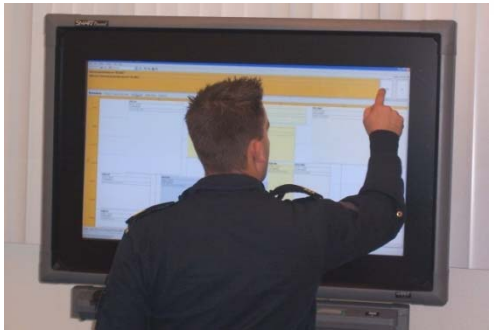


Control room

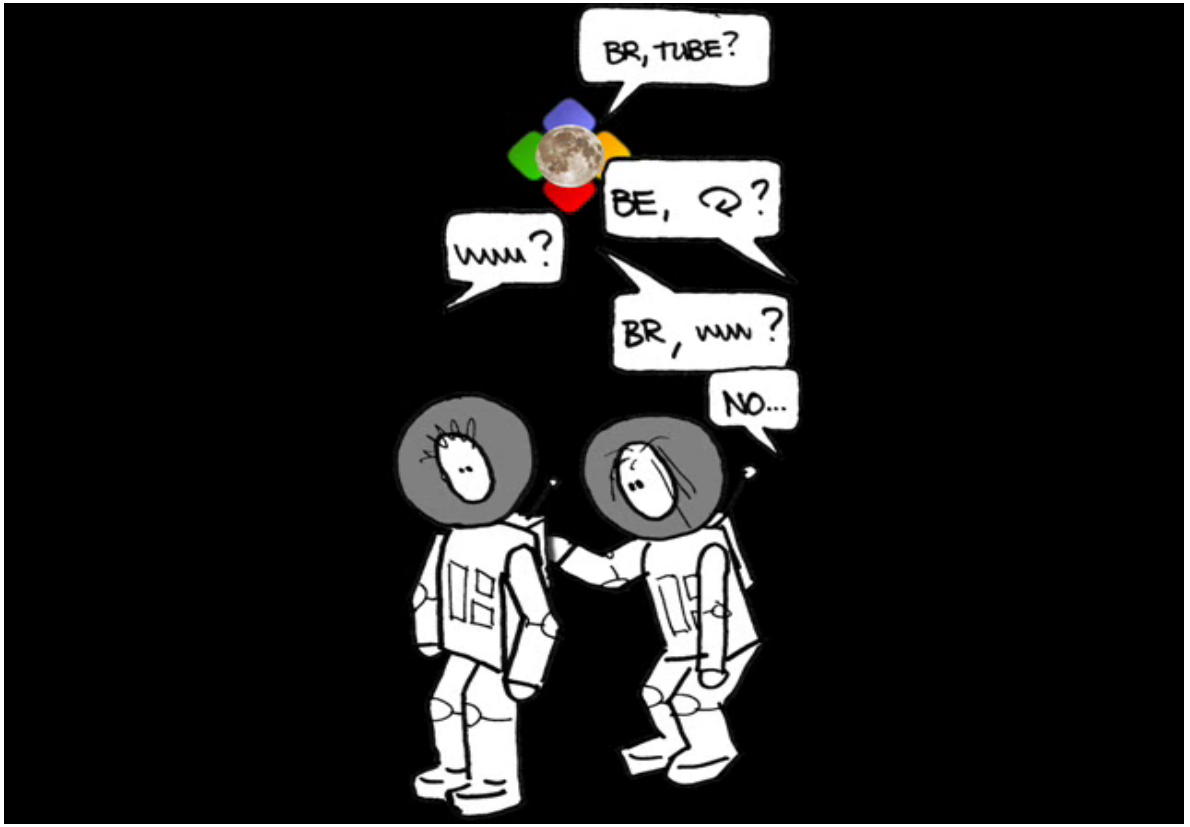


Benny/Brenda

Herman



Evaluation session



Results simulation based evaluation

- Claims
  - MECA enhanced *effectiveness and efficiency*
    - Would like to have more explicit explanations of reasoning of MECA
  - Participants were *satisfied* with MECA
    - Would like to have higher feeling of control
  - MECA was *easy-to-learn*
    - Participants knew where to find relevant information
    - Support should be better integrated in task execution process and context of operation

Results simulation based evaluation

- Claims
  - MECA improved *situation awareness*
    - Adding predictive displays
    - Context dependent situation awareness support
  - *Trust* varied as expected in the scenario
    - Hard to measure in one day session
  - *Emotion*
    - MECA should adapt feedback according to urgency of situation

Results simulation based evaluation

- Scenario
  - Well suited to test MECA
  - Suggested improvements to the scenario for a next test:
    - Voice communication between astronauts
    - Include nominal situations to a larger extent
    - Test MECA over a longer period
    - Feeling of control (by means of a branching scenario)

- Prototype
  - Allow more flexibility in next evaluation (too rigid)

Results simulation based evaluation

- User interface
  - differentiate warnings for nominal situations and anomalies

- Cognitive task load and emotion
  - Participants were positive about MECA measuring cognitive task load and emotion
  - MECA should monitor task load and emotion automatically
  - Be able to turn task load and emotion measuring functionality on or off

Results simulation based evaluation

- Evaluation led to confirmation of existing requirements
- On some requirements we received no feedback
  - Interface requirements and more technical requirements
- Some requirements were not present in prototype evaluation
- Adding of new requirements

Overall conclusions

- Virtual environments
  - Mobile devices in a dynamical context
  - Easy logging of events
  - Easy to extract data and process it into useful visualizations to interpret the results.
  - Participants liked the experiment and were motivated
- Validated? In a simulated environment:
  - More realistic performance
  - “Realistic” subjective emotion after critical event
  - More intense feeling of presence
  - Better situation awareness

- We are looking for 3 master students!
- Terrestrial test with MECA
- End 2008/ begin 2009

- More information
  - Nanja.Smets@tno.nl
  - www.CrewAssistant.com
  - On Wednesday: HB 10.130