

IAC-12-A1.1.23

IMPROVING CREW SUPPORT METHODS IN HUMAN-MACHINE TEAMS FOR LONG-DURATIONS MISSIONS

Nanja J.J.M. Smets
TNO & Delft University of Technology, Netherlands, nanja.smets@tno.nl
Iris Cohen[1,2], Mark A. Neerincx[1,2] , Willem-Paul Brinkman[2] and Jurriaan van Diggelen[1]
[1]TNO, [2]Delft University of Technology, Netherlands,
iris.cohen@tno.nl, mark.neerincx@tno.nl, w.p.brinkman@tudelft.nl, Jurriaan.vandiggelen@tno.nl

We evaluated Mission Execution Crew Assistant's (MECA) crew support methods during the MARS-500 experiment (520 days). MARS-500 provided a unique test platform, because of its setting, where a small crew is isolated for a long duration simulating a manned Mars mission. Thus more prolonged or repeated usage of MECA could be tested. The evaluation focused on core support functions that concern prolonged or repeated usage of MECA.

For the MECA experiment, two groups of three astronauts trained and gamed every other week for thirty minutes (including procedure training and entertainment gaming). The astronauts communicated via chat. MECA collected information on crew condition (social network, Emotional State) and performance (effectiveness and efficiency of operations during training and gaming), and provided (simple) feedback on crew condition and performance. Research questions were:

- How to record and interpret social, cognitive and affective processes during computer based tasks?
- How to provide individual and team feedback on these processes?
- How the crew responds to such ePartner (electronic Partner) actions?

The Mars500 support functions were perceived useful in general, but particular improvements in the content, personalization, usability and attractiveness were needed to establish high performance profits and end-user acceptance. Team-member's inclination to express Emotional State changes differed consistently between the two groups with different cultural and social characteristics. Memory deficiencies differed for the crew members, providing important requirements for MECA support. For better effects on performance, user modeling methods should be applied to tailor the support functions to the individual situated support needs. Providing a free-format text communication tool (like chat) offers major opportunities to collect data on Emotional States and group cohesion.

Important lessons learned were: richer content and interactions are needed for long duration, empirical studies of this kind. Lack of a common language brings additional constraints and costs. Constraining or stripping game functionality to control user behavior had a negative effect on user motivation. The prototype and test set-up should induce an adequate level of intrinsic motivation. The crew-members liked to have timeline support. Large size and diversity of data; proving to have much potential to monitor and interpret crew(-member) conditions, performances and perceptions.

I. INTRODUCTION

Long-duration missions (e.g. to the Moon, asteroids, or Mars) require astronauts to collaborate and interact with complex computerized equipment and facilities under dynamic and hazardous conditions. The Mission Execution Crew Assistant (MECA) comprises crew support that acts in this ubiquitous computing environment as an "electronic partner" (ePartner), helping the crew to take care of their mental and social conditions, to train and schedule tasks during nominal and off-nominal situations, and to enhance the shared situation awareness (SA), sense making, and problem solving processes during operations. For the development of MECA, we apply a human-, task- and context-driven design and evaluation approach [1,2].

Framework

Figure 1 illustrates the core high-level MECA functions that were studied in the MARS500 experiments. MECA's activity monitoring and scheduling support proceeds in four stages, starting in the upper left corner:

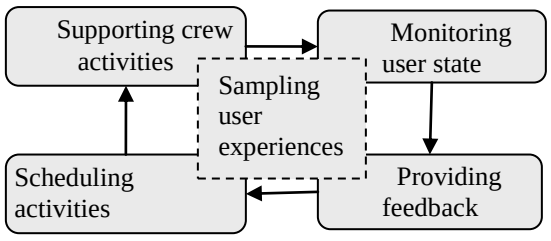

Figure 1: High-level MECA functions.

1. **Supporting crew activities:** MECA is always present. In this sense MECA provides continuous support (e.g., informing team-members on the state of resources that are involved in the current operation). However, for some activities, MECA is latently present and only becomes active in off-nominal situations. For other activities, MECA plays a more prominent role and provides the framework itself within which these activities take place.
2. **Monitoring user state**: MECA monitors the crew members' performed activities. This includes assessing what the crew members are doing, measuring their performance, and assessing their cognitive state (i.e. their Emotional State and task load).
3. **Providing feedback:** MECA provides feedback to their users by presenting the monitored data. This allows a user to better understand his current task behaviour in the context of his performance over time.
4. **Scheduling activities:** MECA supports activity scheduling by offering an electronic timeline tool, and by automatically scheduling (or suggesting scheduling) tasks, depending on its interpretation of the data on previous activities.
5. **Sampling user experiences:** MECA mediates the maintenance and appraisal of memorable experiences and events with a multimedia annotation tool. This function should support the astronauts to reflect on previous activities and happenings in a constructive way to improve resilience (cherishing of successes, coping with stressful events and learning).

Each of these high-level core functions can be implemented at different levels of sophistication. On the one hand, we can think of a full-fledged Artificial Intelligence (AI) implementation, where MECA automatically "reads" the user's mind to assess the user state, provides high level feedback by saying things like "take it easy", "come on", and "don't worry", and by acting as a smart secretary by appropriately (re-)scheduling the user's planning. Going a step further would be to give the user advice on how the reduce stress levels or to appraise the situation in different way to again change someone's emotion, mood or stress level On the other hand, we can think of a simple implementation, where the system monitors the user state by collecting questionnaire data, where the feedback consists of a statistical interpretation of this data, and where the system helps the user to reschedule the timeline by presenting relevant information.

Because the full-fledged AI implementation is currently still science fiction, MECA's current implementation is closer to the simple implementation than to the AI implementation. Nevertheless, we believe that also the principles behind more advanced support systems can be explored by evaluating more simple prototypes. Incrementally refining and adding functions, the 'simple' version will evolve, step-by-step, into a more intelligent version [3],[4].

Experiment

In Mars-500 six crew candidates (three Russian, two European and one Chinese) were sealed in isolation from June 2010 until November 2011. In total they were isolated 520 days, in which they were on a simulated trip to Mars, including the journey to Mars, landing on the planet and a return journey to Earth. The participants had contact with each other and voice contact with a simulated control centre and family and friends. And all contact experienced a simulated increasing delay till 20 minutes [5].

MARS-500 provided a unique test platform for MECA, because of its setting in which a small crew is isolated for a long duration to simulate a manned Mars mission. In this setting, more prolonged and repeated usage of MECA could be tested.

III. METHOD

III.I PARTICIPANTS

In total, there were six participants. The participants were part of the Mars-500 program by ESA and IBMP [6]. For the tasks it was necessary to divide the six participants in two groups of three persons. This also led to a logical division of one proficient English speaking group and one Russian speaking group (who were less proficient in English).

III.II EXPERIMENTAL DESIGN

The participants performed a pre-isolation session, this was an instruction and training that took one day. Then from the start of the isolation period they performed the experiment once every two weeks for half an hour, with exception of the simulated Mars landing (which made the participants miss two sessions).

The participants were not allowed to talk with each other during the session, they were only allowed to use the chat functionality provided.

III.III TASK

The participants started by logging into the system with a user name after which the participant filled in a general questionnaire. This questionnaire was used to see if the data of this session was not corrupted by external factors (e.g. the lack of sleep caused by isolation).

After this questionnaire they were shown a web cam viewer. The participant had to make sure the webcam

was set-up correctly. After this, they were shown an overview screen, a chat client, and a timer.

The overview screen (see Figure 1) showed a timeline from which the different games could be started, and showed the average performance per game of the last sessions. Before and after playing a game, the game performance screen was shown which displayed statistics on the performance of previously played games. They played three games per session in the following order: Colored Trails (CT), Collaborative Trainer (COLT) and Lunar Lander (LL).

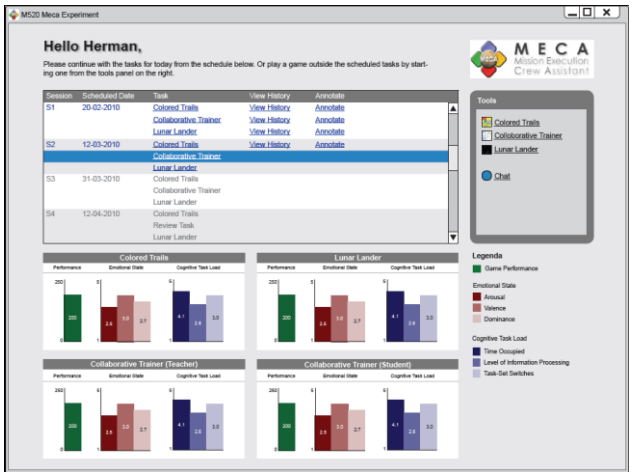

Figure 1: Overview screen, with the timeline, performance graphs and direct links to the applications and games.

Before and after the participant played a game, the game performance screen was shown (see Figure 2). This ensures that the participant frequently sees the information on his past performance, Emotional State, etc. The history viewer showed the following information:

General:

- Emotional State, was represented by a 5-point Likert scale. The Emotional State questionnaire consisted of three scales [7].: arousal (1-5), valence (-2 - 2) and dominance (1-5) The participants had to fill in their current emotion
- Cognitive Task Load, by representing the Time Occupied, Level of Information Processing and Task-Set Switches values from the CTL-questionnaire [8].

Performance (graphs)

- Lunar Lander score
- COLT Teacher score, based on evaluation by students
- COLT Student score, based on the quiz results and the evaluation by the teacher.
- Colored Trails results

From the graphs the participants had an overview of performance, CTL and ES during Mars-500.

The game performance screen allowed the participants to annotate his or her sessions by adding texts, photos, or audio segments. The annotation viewer was used to access specific information of a task that had been executed. This supports the memory of the participants.

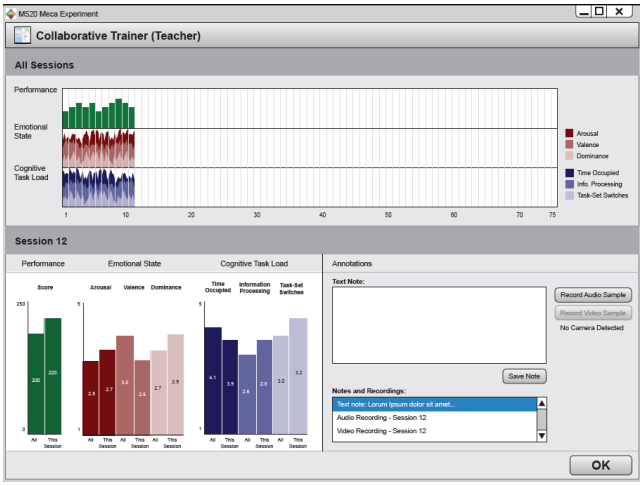

Figure 2: Game performance screen, with the current scores (lower left corner), all available mean scores per session (upper half) and the annotations space (lower right corner)

Collaborative trainer

Collaborative trainer is a chat-based procedure trainer with alternating teacher/student roles. The COLT sessions were manipulated in the following ways:

- Role: Teacher or Student
- Payload : Water Tank or Cardiopres

The water tank was a simulated task of managing a small water filtration unit. The procedures are learned during the regular COLT sessions; the participants had to use this knowledge to keep the water tank operational.

The Cardiopres simulator is a medical device that helps astronauts with medical tasks. The participants had to use the simulated Cardiopres during the COLT-student task. They can fill in different medical information about a fictitious patient.

Hence, in total there were six different configurations. Every session the teacher changed. The first 25 weeks, every first two sessions taught the "Cardiopres" payload and the "Water Tank" was taught in the third session. The water tank scenario was always taught by the same teacher. This allowed for the MECA team to give better instruction to the teacher.

Every session the students filled in a test, concerning the procedure learned. This questionnaire contained

questions regarding facts and procedures. The questions were used to test retention of this knowledge.

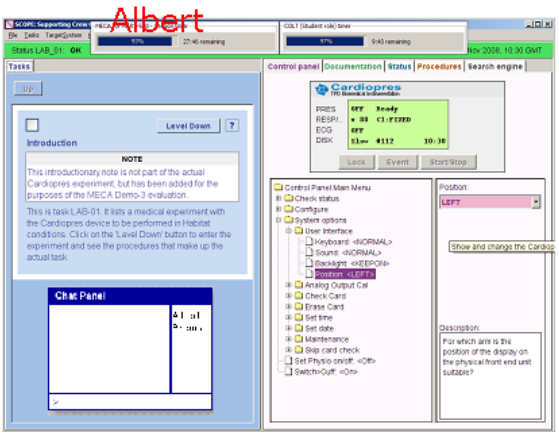

Figure 3: Screenshot with part of the COLT-Student screen, with the chat screen in the left lower corner.

Memory and review task

Instead of the COLT Water Tank or Cardiopres, the participants performed a memory task during four of their sessions (session 12, 17, 26 and 31) and a review task during two of their sessions (session 13 and 32).

In the memory task, they were asked about a specific session, what their role (student or teacher) was, if aspecial event occurred and if their valence in that task differed from their mean valence. This event could be a procedure with the water tank or the learning of a Cardiopres procedure. The participant was asked to give an indication of when this event had happened and what exactly happened. We expect that MECA supports their SA on these aspects in the overview and game performance screen.

The review task was used to get an early evaluation on the ePartner notion and MECA and to see if the participants like the general idea of MECA.

Colored trails

Colored Trails is a negotiation game [9]. It is played on a rectangular board consisting of squares, colored in one of several predefined colors. Each player possesses a piece located on the board and a set of colored chips. A colored chip can be used to move a player's piece to an adjacent square (diagonal movement is not allowed) of the same color. The general goal is to position pieces onto or as close as possible to a goal location indicated by a flag. Although, there is a single goal (flag), each player receives points purely based on its own performance. Figure 4 shows an example of the board, goal (indicated by the flag) and player locations (indicated by P1, P1 and R).

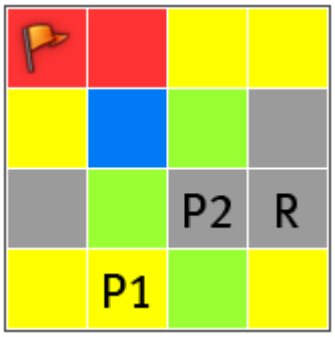

Figure 4: The game board representation

Lunar Lander

Lunar Lander is a fun small game where the player has to land a Lunar Lander on the Moon, see Figure 5.

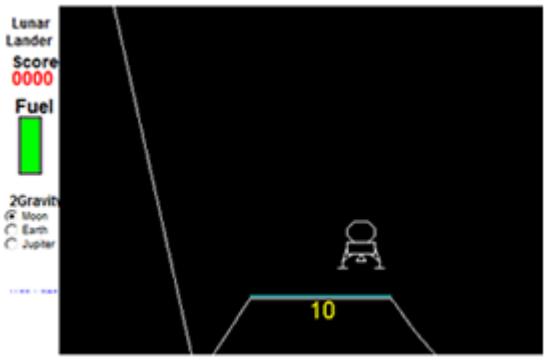

Figure 5: Screenshot of the Lunar Lander game.

III.IV MATERIAL AND SOFTWARE

The participants had a workplace available for their laptop and should start the laptop and MECA software. The participants were in different parts of the M500-station. They had a place for their laptop and were connected to a power outlet. The laptops were connected to each other via Wi-Fi.

The laptops were of the type: ACER Aspire 7715, with an integrated webcam. The power supplies were adapted to be able to fit the Russian power sockets.

The webcam window could be used to direct the webcam such that the face could be properly recorded. The webcam window showed a live image of the webcam, regardless of whether it was recording or not. The recording and pausing options were not accessible to the participant. The recording times were specified per session.

The chat functionality that was provided was peer to peer. There was a possibility to broadcast messages, which was implemented by automatically sending a message to all persons.

IV. RESULTS

We will first discuss the results from the Cognitive Task Load and Emotional State (ES). We will then have a look at the results from the memory task and the review task.

Cognitive task load and Emotional State

First, the subjective measures of Cognitive Task Load and Emotional State were displayed visually per participant and per task in graphs. This shows the progress of the ES and CTL over the whole mission. A first look at these graphs show a clear difference of the subjective scores between the two teams. Figure 6 shows data of a participant of the Russian speaking team, and figure 7 shows data from a participant of the English speaking team. Both graphs show all the sessions of one task. Not every task was conducted every session so the session numbers are not the same for every participant and task. Figure 6 and 7 illustrate clearly that the participant in the English team gave more varying answers to the emotion and Cognitive Task Load questionnaires than the participant of the Russian speaking team. This suggests that answers to emotional questionnaires are dependent of the participants cultural or social background and that the Russian group probably expresses differences in Emotional States less. We continued with analyzing cognitive task load and Emotional State data from the three participants in the English team.

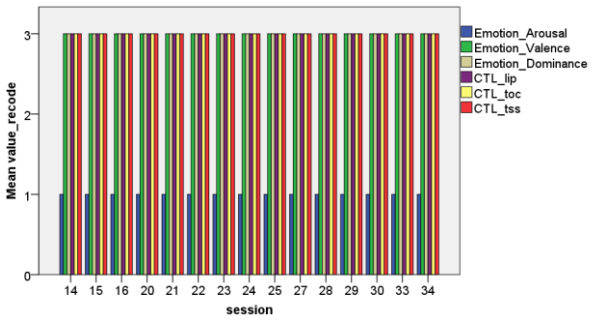

Figure 6: CTL and ES scores from a participant of the Russian speaking team for one task.

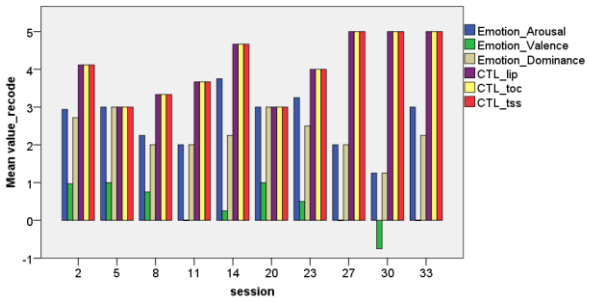

Figure 7: CTL and ES scores from a participant of the English speaking team for one task

Five graphs like figure 6 and 7, were made for each of the participant, showing the ES and CTL scores. To see if ES and CTL influences the results of the task, a table was created that shows interesting sessions (very high or low CTL or ES scores or negative valence) and the corresponding result of the task. A small section of this table can be seen in table 1. Every column shows a task, the rows indicate session numbers per participant. The results of these interesting sessions are shown under the heading 'result'.

Participant team1	**COLT CP**		**COLT WT**		**COLT Teacher**	
	Session	Result	**Session**	Result	**Session**	Result
1	**21**	5	**14**	5	**1**	2
	24	5	**23**	-	**16**	5
	28	5			**22**	5
2	**25**	5	**30**	5	**24**	5
	29	5	**11**	5	**28**	5
3	**1**	4			**5**	4
	15	1			**27**	5
	25	5			**30**	5
					33	5

Table 1. Sessions numbers (in bold) and the corresponding results of the tasks per participant. CP = Cardiopres, WT = Water Tank and T = Teacher.

This table was a guideline for choosing which session the video data would be interesting to analyze with the FaceReader software from Noldus [10]. For participant 1, session 22 (COLT teacher and Lunar Lander) was analyzed. During both tasks the participant showed mostly a sad facial expression with some neutral expressions in between. For participant 2, session 24 was analyzed. In this session, the participant showed mostly surprised, sad or disgusted facial expressions. For participant 3, the video of session 25 was analyzed for the COLT Cardiopres task. The data showed mostly a neutral face, with some sad and disgust. So, there was consistency between the data from the table and the Emotional State assessments of the FaceReader.

Another helpful way to interpret the ES data, was to examine only the scores given at the beginning of each session, before any of the tasks had started. These scores should show the overall Emotional State of the participants independent of the task. Only one participant of team 1 showed interesting ES and CTL scores and negative valence at the beginning of a few sessions. It would be useful to investigate if an external factor was responsible for the negative mood of this participant. The other two participants show neutral to slightly positive ES data and average CTL scores.

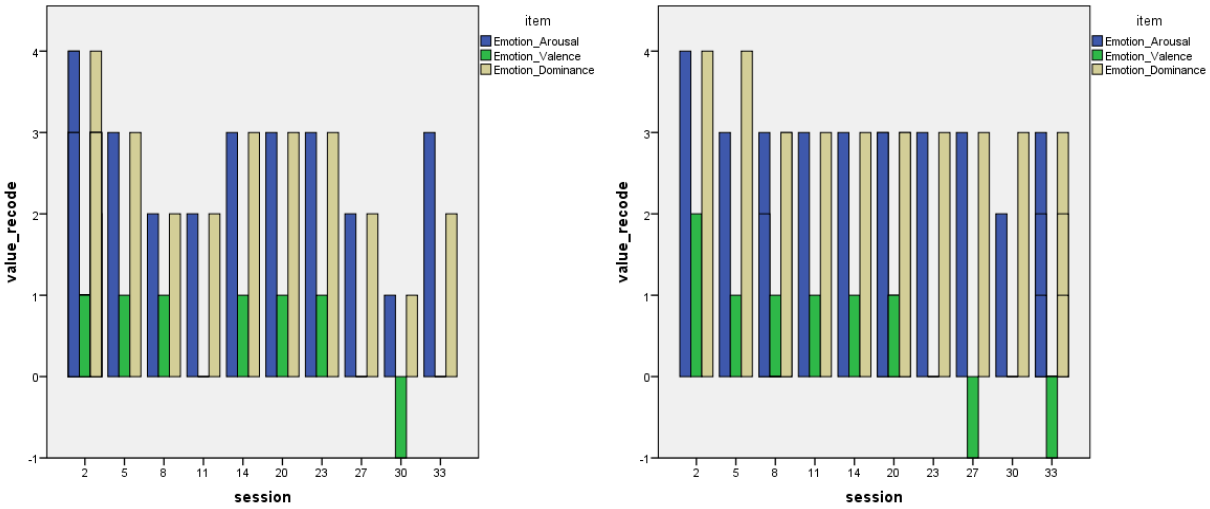

Figure 8: Two participants both scoring 0 or negative on valence for session 27, 30 and 33.

In figure 8, two participants both show negative or zero valence prior to the same task. In this case, it might be helpful to examine external factor that might have caused this, but also interaction during the previous task. Different factors could explain why two participants both showed negative valence before the same session and task.

Memory task

The results from the memory task are discussed for participant 1, 2 and 3. See Table 2 for a schematic overview of the results.

Participant	memory task nr.	Task	Role	Event	Valence
1	1	+	+	+	+
	2	-	+	+	-
	3	-	+	-	+
	4	-	+	+	-
2	1	-	+	+	+
	2	-	-	+	-
	3	-	-	-	+
	4	-	+	+	-
3	1	+	+	+	+
	2	+	+	+	+
	3	-	+	-	-
	4	-	+	+	-

Table 2: This table shows if the answers of participants 1-3 in the memory task were correct (+) or incorrect (-). If they remembered the task, their role correct, if an event occurred and if their valence in the reviewed session was higher or lower than the mean valence.

They had to indicate their task (Cardiopres or Water Tank), role (student or teacher), if and what kind of event occurred and if their valence in the reviewed session was higher or lower than the mean valence.

In all review tasks participant 1 and 3 remembered their role correctly. From the answers it was clear that participant 3 found the task really difficult, because he had a lot more incorrect answers and these incorrect answers were also nonsense answers.

Review task

The participants filled in a review questionnaire in session 13 and 32 where they were asked what they thought of MECA. They had to rate a number of statements on a 5-point Likert scale (1 = disagree to 5 = agree). Below are the statements:

1. I understood the rationale behind the MECA experiment.
2. I enjoyed performing the MECA experiment.
3. The COLT application helps you to learn new procedures effectively?
4. The COLT application provides a convenient environment to train new procedures?
5. Individual performance, cognitive task load and emotional state should be monitored during long duration missions.
6. Individual performance, cognitive task load and emotional state should be monitored automatically instead of using questionnaires.
7. The provision of statistical information on performance, cognitive task load and emotion was helpful.
8. Seeing the relations between performance, cognitive task load and emotion was helpful.
9. The annotation function of the timeline tool was useful.

10. The timeline tool provides a useful start and overview of the activities

The answers to the statements can be found (displayed per group: English or Russian speaking) in Figure 9, 10, 11 and 12.

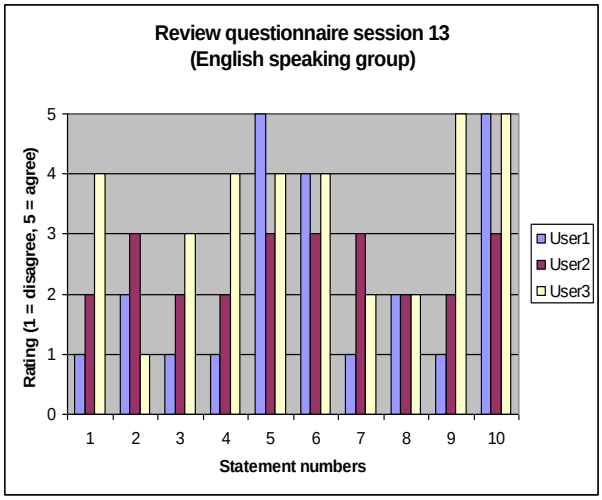

Figure 9: Answers of participants 1 to 3 to the statements of the review task in session 13.

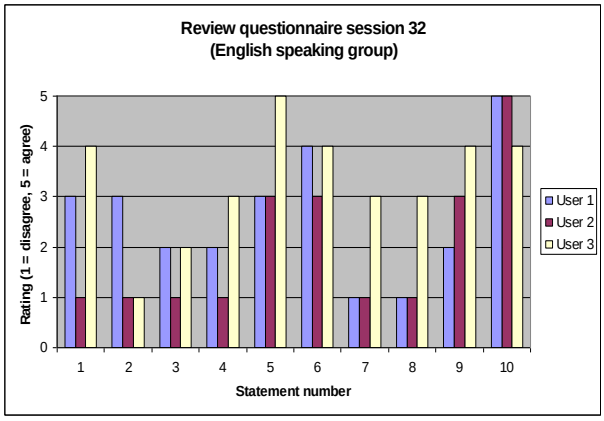

Figure 10: Answers of participants 1 to 3 to statements of the review task in session 32.

In Figure 9 and 10 give a mixed view on the results to the statements Some initial observation is that the English speaking group were positive about the collection of CTL and ES data and they would like it if this was collected automatically (statement 5 and 6). Furthermore, they found the timeline tool very useful in both review sessions (statement 10).

What immediately stands out in Figure 11 and 12, is that participant 6 has not performed the review task in session 32. And that participant 5 gave less varying answers to the statements. Also the Russian group was positive about the automatic collection of CTL and ES data (statement 6) and they agree that it should be collected (statement 5). Another difference is that they were rather negative about the timeline tool, but that this may come from the fact that the Russian group only performed the COLT- water tank and thus had less variation in their schedule. In Figure 11, the answer to statement 9 by participant 6 stands out, this participant was very positive about the annotation tool.

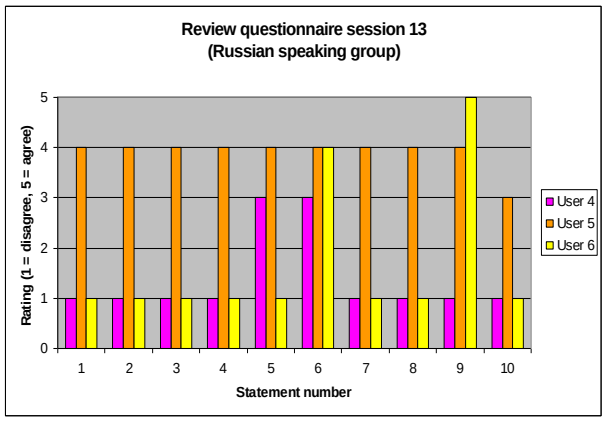

Figure 11: Answers of participants 4 to 6 to statements of the review task in session 13.

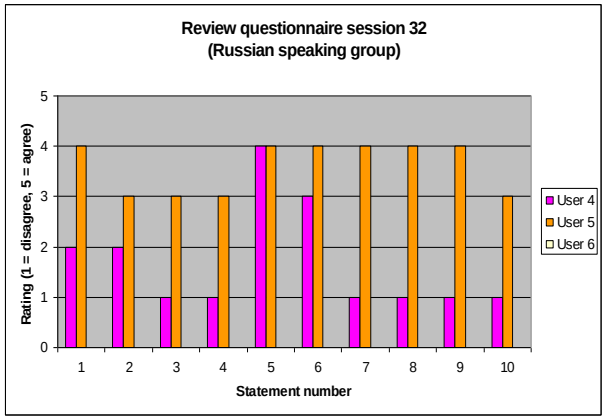

Figure 12: Answers of participants 4 to 6 to statements of the review task in session 32, participant 6 did not perform the review task for session 32.

The review task also encompassed a number of open questions. We will only discuss the answers of the English speaking group to the open questions, the Russian speaking group did not answer the open questions apart from the note that MECA should be translated. The English speaking group gave some feedback on usability (e.g. too many submenus) and on the games. They found CT boring and would have liked it if COLT was applied to an actual payload.

V. DISCUSSION

Cognitive Task Load and Emotional State

The first analyses of the Emotional State and Cognitive Task Load data shows that there is a difference in the way the participants answered the questions. The team with the participants with a Russian background gave less varying answers and mostly gave a 'neutral state' with their answers. This is the reason why this paper focusses mainly on the data of the English speaking team. For a later phase, data from both teams will be analysed in more detail thereby considering potential explaining factors. Besides the cultural difference in Emotional State and Cognitive Task Load, there were some more result that stood out in the data.

Video data from a few sessions in which the participants replied to feel negative valence, were analyzed with the FaceReader software from Noldus. Because of poor lighting conditions and varying postures of the participants, the analysis led to incomplete data for some sessions. The analysis showed that the facial expressions of the participants consisted mostly of negative emotions. This data must be compared to video data from sessions in which the participants said to have a positive valence, to see how accurate the FaceReader analysis is.

We also looked at the Emotional State at the beginning of the sessions. This should show the Emotional State of the participants independent of the task, because they have not occurred yet. For only one participant this analyses showed interesting data. Finding out why this participant felt differently from the other participants in the same environment will be one of our next actions.

To look at the effect of isolation, it is interesting to analyze how far Emotional States between users are aligned. They, off course, share the same environment and have to work together in most of the tasks. Some data showed that in some sessions, more than one participant reflected to feel the same negative valence. This effect is important for long duration and isolation missions if it means that negative Emotional State of one participant will reflect on other participants. The reason for this effect we have found should be looked into.

Review Task

The review task gave mixed results, but does give some interesting feedback on MECA. The timeline tool was agreed to be useful by the English speaking group and not the Russian speaking group. Furthermore, all participants indicated that CTL and ES should be monitored (preferably automatically). Richer content and interactions are needed for long duration, empirical studies of this kind. Constraining or stripping game functionality to control user behavior had a negative effect on user motivation. The prototype and test set-up should induce an adequate level of intrinsic motivation.

Memory Task

We expected that by showing the current and past performance, ES, CTL and timeline that this would support the participant's SA about these aspects. The results of the memory task showed differences amongst the participants. From the data it was clear that participant 2 found it difficult to fill in the memory task (a lot of mistakes and filled in nonsense data). Participants 2 and 3 performed better on the memory task. Especially roles an events, which were shown in the timeline tool. This corresponds to the outcome in the review task that they found the timeline tool useful.

VII. CONCLUSION

The Mars500 experiment had some major new test-conditions:

1. The MECA prototype was evaluated for a really prolonged period for the first time. Such experiments take place seldom, due to several reasons such as the costs, the availability of end-participant representatives, the risks for organizational and/or software failures and the constraints on planning. The software and test protocol proved to be robust, providing the data collection aimed for.
2. Intensive team work was required in COLT and Colored Trails, setting high demands on collaboration. The new use cases addressed these collaboration demands, leading to the specification of a new requirement for collaboration support.
3. The test encompassed an extensive simulation of crew members who have to perform space operations in isolation. The software recorded their behavior consistently.

These new test-conditions gave an extensive amount of data that was used to address the questions mentioned in the abstract. In this experiment social, cognitive and affective processes were recorded and interpreted during computer based tasks. Team-member's inclination to express Emotional State changes, differed consistently between the two groups with different cultural and social characteristics. The FaceReader as a method to monitor these states unobtrusively and automatically proved to provide consistent results, but lacking the robustness to collect reliable data in dynamic contexts (such as lighting and posture conditions)

Feedback on individual (e.g. overview of past valence) and team (e.g. feedback on results of the COLT task) processes were given to the participant but did not support the memory of the participants.

The crew members response to the MECA ePartner indicates the importance of an overview in the timeline and that they were positive about showing and automatically measuring Cognitive Task load and Emotional State. All these results provide important requirements for MECA support.

Acknowledgements

Part of this research was conducted for the MECA project. This is a development funded by the European Space Agency (contract numbers 19149/05/NL/JA and 21947/08/NL/ST).

Part of this research was conducted for project (056-22-010) and is supported by the Dutch FES programme "Brain and Cognition: Societal Innovation".

REFERENCES

1. Neerincx, M.A. & Lindenberg, J. (2008). Situated cognitive engineering for complex task environments. In: Schraagen, J.M.C., Militello, L., Ormerod, T., & Lipshitz, R. (Eds). Naturalistic Decision Making and Macrocognition (pp. 373-390). Aldershot, UK: Ashgate Publishing Limited.
2. M.A. Neerincx, A. Bos, A. Olmedo-Soler, U. Brauer, L. Breebaart, N. Smets, J. Lindenberg, T. Grant, M.Wolff, The Mission Execution Crew Assistant: improving human-machine team resilience for long duration missions, in: Proceedings of the 59th International Astronautical Congress (IAC2008), 2008, 12pp. Paris, France, IAF, DVD, ISSN 1995-6258.
3. Neerincx, M.A. and Grant, T. (2010). Evolution of Electronic Partners: Human-Automation Operations and ePartners During Planetary Missions. Journal of Cosmology. Vol 12, 3825-3833
4. Neerincx, M.A. (2011). Situated Cognitive Engineering for Crew Support in Space. Personal and Ubiquitous Computing. Volume 15, Issue 5, Page 445-456
5. ESA Website, last accessed: 23 August 2012. http://www.esa.int/SPECIALS/Mars500/index.html
6. Russian Academy of Sciences' Institute of Biomedical Problems. The MARS-500 isolation experiment. http://mars500.imbp.ru/index.html. Accessed 10 September 2012.
7. Zijlstra, F.R.H., (1993) Efficiency in Work Behaviour: A Design Approach for Modern Tools, Delft University Press, Delft, 1993.
8. M. A. Neerincx, (2003). “Cognitive task load design: model, methods and examples,” in Handbook of Cognitive Task Design, E. Hollnagel, Ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2003, ch. 13, pp. 283– 305.
9. Grosz, B. J., Kraus, S., Talman, S., Stossel, B., and Havlin, M. (2004). The influence of social dependencies on decision-making. initial investigations with a new game. Proceedings of the 3rd international joint conference on autonomous agents and multi agent systems, 2:782–789
10. Noldus. FaceReader product page: http://www.noldus.com/human-behavior-research/products/facereader. Accessed 10 September 2012.