DR 7.3: Scenario-based evaluation with multiple robots

Joachim de Greeff[*], Nanja Smets[†], Rainer Worst[‡], Mario Gianni[§], Norbert Pahlke[¶] and the TRADR consortium

[*]*Technische Universiteit Delft*
[†]*Netherlands Organization for Applied Scientific Research*
[‡]*Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS*
[§]*Alcor Laboratory, Department of Computer, Control, and Management Engineering "Antonio Ruberti" - Sapienza University of Rome*
[¶]*Institut für Feuerwehr und Rettungstechnologie FDDo*

⟨`J.deGreeff@tudelft.nl`⟩

Project, project Id:	EU FP7 TRADR / ICT-60963
Project start date:	Nov 1 2013 (50 months)
Due date of deliverable:	Month 38
Actual submission date:	XXX 2017
Lead partner:	TUD
Revision:	February 23, 2017
Dissemination level:	PU

We report on progress achieved in Year 3 of the TRADR project in WP7: *Scenario based evaluation with multiple robots.* We discuss the further refinement of the use cases for Year 3, describe discussions regarding the TRADR unit that took place this year and report on the organization and execution of the Year 3 exercises. Additionally, a description of the setup and launch of the TRADR system as used during T-Eval 2016 is provided.

Executive Summary

This report documents milestone MS7.3: scenario-based evaluation Year 3 for WP7. This milestone consists of task T7.3: to perform a *Scenario-based evaluation with multiple robots*. Building on the foundations developed during Year 1 and Year 2, we further refined the scenario and use cases, which were subsequently used during two evaluation exercises (T-JEx and T-Eval) held in July in Prague and in October/November in Dortmund, Germany respectively.

Role of Scenario-based evaluation in TRADR

For Year 3, the goal of this workpackage is to perform a scenario-based evaluation with multiple robots. Towards this end, the user-needs, the scenario and the use cases were refined for Year 3. The TRADR Joint Exercise (T-JEx) serves as an exploratory field study during which new components, modules and use cases are tested, while the TRADR Evaluation (T-Eval) is a more formal evaluation with less emphasis on exploration and more on maturation of the system. A high level of end-user involvement remains important in TRADR; as such there is tight collaboration with the end-users in the creation of the scenarios and use cases to ensure realism and fidelity.

Contribution to the TRADR scenarios and prototypes

This workpackage is responsible for defining and creating the scenario and use cases. The use cases have been adapted to account for important aspects in Year 3, which are *multi-robot collaboration*, *multi-sortie 3D model fusion* and an *increase in complexity* of the exercise. In this deliverable we report on the formulation of the TRADR scenario and use cases, and the evaluation of various components and prototypes as part of a larger integrated scenario. By embedding various components and prototypes developed by the TRADR partners into an integrated TRADR system, and evaluating this in a field experiment, we gain valuable insights in how these components perform as part of a greater system, rather than in isolation. Following the methodology developed last year, the exercises allow us to take stock of the performance of the TRADR system on a holistic level, while simultaneously interdependency between various system components can be assessed. This experience helps to formulate targets and priorities for the next year.

Persistence

WP7 captures the notion of persistence by formulating use cases that put demands on the system in such a way that persistence is required to perform well. This is reflected in the use cases (details of which can be found in Annex [1] (Annex Overview 2.1)). For instance, use case #7 and #8 address the systems ability to build on previously recorded maps and use these for navigation and planning, use case #12 addresses UGV patrolling which is aimed at monitoring a particular area for a prolonged period of time, various use cases deal with the detection of victims, fires, smoke and other points of interest, and use case #18 describes how the reporting tool can aid in the transition of knowledge and situation awareness to a new rescue unit. The scenario is formulated in such a way that an adequate response requires accumulation of data and situation awareness through multiple sorties over potentially multiple days.

1 Tasks, objectives, results

1.1 Planned work

The following work was planned for Year 3.

- User needs analysis: it needs to be analyzed what the user needs are when multiple robots are in the field during an individual mission.
- Refinement of the socio-technical design rationale.
- Planning of scenario-based evaluation: based on the user needs analysis, a scenario needs be defined that adheres to the multiple robots during an individual mission focus of this year.
- Define methods and metrics: methods and metrics need to be adapted to multiple robots in the field during individual mission, meaning that we want to measure how the team uses information from others and if they make requests to others.
- Assess with end-users: we include larger fires which initially will act as (dynamic) barriers to exploration. Spatial complexity is increased by including both above- and under-ground exploration (e.g. tunnels, sewers, collapsed structures). Besides the temporal dynamics of smaller and larger fires, we also extend the scope of the response to between 2 and 4 days.

1.2 Addressing reviewers' comments

1. **Multi-robot collaboration and human-robot teaming should be represented more explicitly in the multi-level evaluation methodology diagram wherein different pillars are elicited in the middle level.**
   Multi-robot collaboration and human-robot teaming is now explicitly represented in the evaluation levels diagram, see Figure 1.

   As multi-robot collaboration is an important topic in Year 3, emphasis was placed on the creation of use cases that involve multi-robot collaboration. Towards this end a number of multi-robot tasks were defined, such as e.g. offline coverage (exploration of a known environment), online coverage (exploration of an unknown environment), patrolling (continually surveying an environment by a group of robots) and sensor network deployment (deployment of robot sensors in order to completely cover a given region. For details, see Annex [1] (Annex Overview 2.1), Section 2.1.

2. **Why the robotic arm on top of one of the UGVs was not used by end users in the evaluation exercise should be clearly explained in the deliverable D7.2. In particular, it should be clarified whether this is caused by a lack of usability or further work should be carried out to make the asset more useful for end-users in future exercises.**
   The fact that the robotic arm was not used by end-users during the Year 2 exercises was due to the open-ended approach of the evaluation exercise. In Year 2 our approach has been that end-users were left relatively free to explore and use the TRADR system as they saw fit. It so happened they concentrated on UGV control and exploration with a focus on gathering situation awareness, and did not proceed to, e.g., picking up samples. This resulted in the arm not being used to its fullest potential. To remedy this situation, we have expanded and clarified the training procedure with respect to arm usage. Additionally, the Year 3 scenarios were relatively less free; picking up a chemical sample using the robot arm was an explicit task which was successfully as part of the T-Eval integrated scenario (see Annex [4] (Annex Overview 2.5)).

3. **It is recommended that the consortium clearly identifies quantitative measures of progress for the GUC and EUC scenarios, so that clear progress can be distinguished from one exercise to the next. Similarly, as cognitive load is a repeating factor, it should be measured or at least assessed qualitatively, using standard evaluation techniques commonly used in scientific circles.**
   Defining the Generic Use Cases (GUCs) and Evaluation Use Cases (EUCs) structure in Year 2 was useful, as it enabled us to think about use cases in a structured manner. It however turned out that the uptake and utilization of a clear formal distinction between the two was a bit limited. For this reason the The GUC/EUC structure was not expanded upon in Year 3. Instead, we formulated the use cases in a more story-like fashion, which allowed for a fine-tuned targeting of the actual functionalities of the TRADR system that the use cases are meant to capture. This is effectively the same level of description as the EUCs used in Year 2. Details on the Year 3 use cases can be found in Annex [1] (Annex Overview 2.1)

   Regarding a systematic quantitative comparison of performance on a specific use cases over the years, this proved to be rather challenging – if not impossible – as there is too much variation in the functionalities, complexity and configuration of the TRADR system going from one exercise to the next. To do a proper comparison of a particular feature one would need a more controlled and static system. Instead, the TRADR system is ever evolving and quite substantially changing a significant portion of core functionality, which as such does not allow for very rigid numerical testing. To evaluate where we are and establish progress, we have used a more qualitative approach in which we describe how parts of the system functioned, what was used during the exercises (T-JEx and T-Eval), what went well, what needs to be improved etc. while also collecting feedback from the end users.

Furthermore, ROMA has been working on a framework for virtual simulation (described in DR4.3 [6]), working towards the creation of a platform for both coupled and decoupled testing (for integration reliability purpose only). Future work on this may include quantitative evaluation also based on results in virtual simulation.

Additionally, Fraunhofer has started monthly test cycles in which a large number of aspects of the system are tested. This allows for a more systematic tracking of which features are working and which are not. These test cycles proved to be very useful, also in preparation for upcoming exercises. A detailed example of how such a report looks is provided in DR6.3, Annex 2.2 [11].

Next to this, to capture the extent with which repeatable benchmark tests are being used by TRADR partners for the respective functionalities that they are working on, we have started collecting these benchmarks in an overview. Examples of these are e.g. a testbed specifically aimed at Adaptive Traversability, speech and NLU tests described in DR.3.3 [8] ("Speech Interaction Processing in TRADR") etc. If suitable and appropriate, these will be made publicly available.

Regarding cognitive load of robot operators and the TeamLeader, this was explicitly measured during the T-Eval integrated exercise, details of which are described in Annex [4] (Annex Overview 2.5).

1.3 Actual work performed

This section describes the actual work performed in Year 3. We first describe how the scenario and use cases were defined for Year 3 (Section 1.3.1), followed by a description of the TRADR unit (the collection of human and robot actors that make up the TRADR team, Section 1.3.2) and the setup of the integrated TRADR system as it was used during T-Eval, in support of running exercises (Section 1.3.3). After that we describe the scenario-based evaluation exercises (T-Eval and T-Jex) that were conducted in Year 3 (Section 1.3.4), which is followed by the End-user reports regarding these exercises and a summary of the end-user discussion from T-Eval (Section 1.3.6).

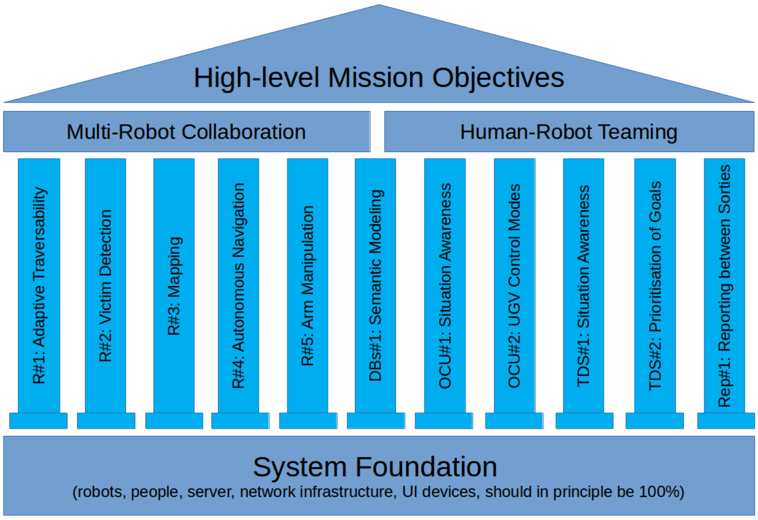

Figure 1: Updated evaluation structure, which explicitly includes multi-robot collaboration and human-robot interaction.

1.3.1 Use cases Year 3

For Year 3, the overall scenario remains the same: a first response mission after a large industrial accident. User needs analysis and the socio-technical design rationale remain largely the same as previous years. following the Cognitive Engineering (sCE) approach [7], we build upon experiences from Year 1 and Year 2 exercises, and expand the system based on end-user feedback. The use cases have been adapted to account for important aspects in Year 3, which are *multi-robot collaboration*, *multi-sortie 3D model fusion* and an *increase in complexity* of the exercise (both the environment and robot- and system capabilities). In collaboration with the TRADR partners, a number of use cases were defined based on the previously mentioned aspects, and by taking into account what is technically achievable. Some notable new use cases are e.g. "UGV control modes (First Person Control/Tank Control)", "Planning based on a map generated during a previous sortie (mapping)", "Map creation in a dynamic environment (mapping)", "Offline and online coverage (multi-robot collaboration)" and "Patrolling (multi-robot collaboration)". For a full overview, see Annex [1] (Annex Overview 2.1).

1.3.2 The TRADR unit

We maintained the view of the TRADR unit as a specialized unit, embedded within a larger first-response organization. In Year 3 some changes in the TRADR unit were considered, the most predominant topic of discussion being the potential introduction of UGV pilots. This role would be akin to the UAV pilot, as to allow a single UGV operator to control multiple UGVs, which in turn are controlled by pilots as intermediates, eventually to be replaced by an automated system. This would have profound implications on the team constitution, effectively shifting the UGV operators' role to a more high-level supervisory one. After an extensive discussion amongst the TRADR consortium partners, it was eventually decided to not introduce this change, but instead retain a team structure similar to the one used in Year 1 and Year 2. A detailed description of this discussion and the considerations can be found in Annex [2] (Annex Overview 2.2).

Furthermore, with the increase of the number of UGVs, it became apparent that a new naming schema was required, as the previously used "UGV Operator #1 = UGV1" did not suffice anymore. To improve both clarity and readiness in communications, a new naming scheme was adopted for the UGVs, in which we explicitly distinguish between UGV operator (denoted as 'UGV OP #x') and UGV robots which are named according to the ICAO phonetic alphabet[1]. For instance, during T-Eval, up to four UGVs were used which were named after the first letter of the respective TRADR partner, thus their ICAO call-signs were `CHARLY` (CTU), `DELTA` (DFKI), `ECHO` (ETHZ) and `TANGO` (TNO).

Another important chance in Year 3 was the decision to drop the human Infield Rescuer. One reason being that this role was either under-utilized or over-utilized, and generally tended to distract the TeamLeader. For instance, during some exercises it was observed that a TeamLeader would totally forget about the Infield Rescuer because he was very much involved in orchestrating UGV actions, while in other exercises the TeamLeader would focus very much on utilizing the Infield Rescuer, as such diminishing the UGVs and UAVs roles. A second reason was that during discussions with the end-users it transpired that they would envisage human rescuers as separate from the TRADR robot team. Thus, the TRADR unit is a specialized unit specifically aimed at information gathering in hazardous environments that are deemed too dangerous for human rescuers to enter. When useful information is gathered and SA has been established through utilization of the TRADR unit, this information can subsequently be passed to a human rescue team when appropriate.

Taking the above into account, in Year 3 the TRADR unit was realized in the following manner.

- Four human team members in remote command post (TeamLeader (TL), UGV OP #1, UGV OP #2, UAV OP).
- Up to two UGVs simultaneously in the field, out of a fleet of potentially 4 UGVs with different payload configurations.

[1] `https://en.wikipedia.org/wiki/NATO_phonetic_alphabet`

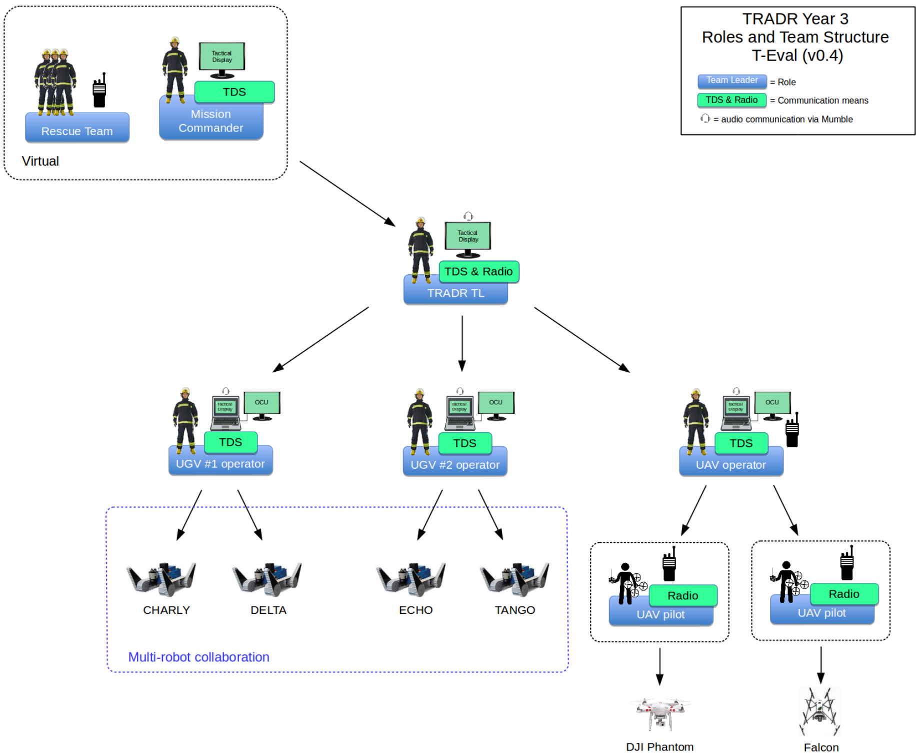

Figure 2: The Year 3 TRADR unit, as actually used during T-Eval.

- Two UAVs (Falcon and DGI), each with a dedicated pilot.

Figure 2 shows a schematic overview of the Year 3 TRADR unit.

1.3.3 Setup of the integrated TRADR system

There is the vision to press a single button to launch the whole TRADR system – however in practice it is not quite as easy. Nevertheless, we made significant progress in Year 3 to define protocols for the setup of the TRADR system and to keep the involved procedures under control.

Particularly during T-Eval, we worked towards making the system setup more organized and streamlined, in such a way that eventually it could be launched by end-users who are not necessarily robotics experts. In this section we describe the work that has been done towards this goal. Note, it is strongly related to WP6 and as such needs to be considered in conjunction with the overall system architecture and organization. However, where DR6.3 is more about the general architecture and system setup, this section relates in detail the status of the system during T-Eval Year 3, and is as such included in this workpackage.

To streamline the setup of the TRADR system, we maintain a Wiki page (see Annex [12] (Annex Overview 2.3)) that describes the setup of the system components according to the phases of the disaster. This starts with the setup of the UAV, the network, the TRADR core, and the TDS workstations. In the next phase the UGVs are connected to the network and configured in a proper way according to their current sensor payload. The Wiki page contains links to subpages where the setup of all components is described in detail.

The most complex part of this process is the setup of the TRADR core, which plays a central role in our system architecture. During T-Eval 2016, we recorded a live audio to explain each single step of it and created a transcript of it later (see Annex [9] (Annex Overview 2.4)). The process is based on the TRADR orchestration tool (cf. section 1.3.2 of DR.6.3).

These contributions can be considered as an intermediate step towards the goal to let the end-users do the entire setup of the system, assuming that it had been already properly integrated as described in DR.6.3. The descriptions provided in the annex are a part of a user's manual to enable persons, who were not necessarily involved in the development of the software, to operate the TRADR system. During T-Eval 2016 this was still achieved by the scientific partners. A next step would be to have a small dedicated group (who are not the developers) set up the TRADR system in a given timeframe, and observe what is needed for non-TRADR people to do this. We will evaluate in the future whether end-users – provided they have received proper training – can do this job or which improvements are needed in order to make this happen.

1.3.4 Scenario-based evaluation exercises: T-JEx and T-Eval

Year 3 featured two field exercises – T-JEx and T-Eval – aimed at integrating the various components developed by the TRADR partners throughout the year and test these as part of an integrated scenario in the field, together with the end-users. While T-JEx has more of an integration and testing nature, T-Eval meant to embody a yearly system evaluation to assess in collaboration with end-users the state of the system.

It was decided to organize T-JEx without external end-users, as to have a stronger focus on integration and technical testing. For T-Eval, the number of end-users was not as high as we would have liked; due to logistic and organizational difficulties only a limited number of external end-users participated in the exercise. Nevertheless, valuable experiences and feedback were collected during both exercises. A detailed report is provided in Annex [4] (Annex Overview 2.5). Additionally, an extended abstract detailing T-Eval was accepted at HRI2017, see Annex [5] (Annex Overview 2.6).

1.3.5 End-user reports on T-Jex/T-Eval

As usual, we asked the end-users to provide a report on the TRADR exercises that were conducted in Year 3, as to obtain their view on TRADR development, the system and its operation. Below this end-user report is provided.

End-user report

Y3 is characterized by a shift on the scenario time-line close to the beginning of an incident. This includes also a step to/into the time critical phase where time is the dominant factor for the injured's survival. Dynamic elements in the scenario complicate the reconnaissance and increase the demand on the task fulfillment for the robots. Therefore the collaboration between them became more important than ever. On the other hand collaboration among the robots and with the humans in the loop opens up a concentration of the different abilities as also acceleration and optimization in the use of time. The Reconnaissance can be done faster now by the collaboration of the humans with the robots especial because the robots are able to drive automatically and react on the dynamic of a complex situation.

The next two subsections deals with the results of the exercises of year 3 and describe the test scenarios as also the challenges and results from end user's point of view.

TRADR Joint Exercise (T-JEx) The TRADR Joint Exercise took place in Prague at the TBA from 04th to 08th of July 2016, Czech Republic, organized by CVUT Prague. Due to the available resources the exercise scenario was simpler but the evaluation focused the different functionalities of the integrated system. The tests were performed by the end users involved in the project. External end users from the fire department of Prague were invited and two of them participated as experts for civil protection learning from and discussing the functionalities of the system.

The scenario itself consisted of two independent action areas with different tasks and challenges. One area offers an obstacle for testing the autonomous flipper control and victim detection, the other one for obstacle avoidance under smoky conditions and gas detection. First dry runs after the system set up offered technical problems, but they could be solved during T-JEx so that a successful test could take place. Some aspects in more detail: the handling of the robot arm was evaluated separately but also during the exercise and could satisfy the expectations. Its use was easy to learn and the functionality could support the operation as shown in the exercise hereafter. The TDS run stable and offered the users features for the operation management. Georeferenced information gathered on a map, an improved design such as additional status information layers led to a better handling of the TDS. An overload of information is prevented without increasing the complexity. With clearing the code database queries became faster which accelerates the performance respectively the situation awareness.

The TDS offers in combination with a reporting functionality a well-rounded operation management tool for a persistence in situational awareness during the several day operation. But the current reporting tool supports the operation management not completely satisfactory. A broader range would be needed to be "the tool". That means the record of spoken traffic and gathered information by sensors is the half and should be complemented by recording the TDS activities. All together would offer the feature of an incident log and support every user also robots keeping the situation awareness current, because it is not only important to know how the situation itself is representing but also how the team reacts on it and tries to get the scenario under control. The incident log would enable an overview of carried out work, task allocation and task planning. The victim detection, working with an infrared camera, supported the operator by reducing the amount of monitored screens and directed its attention to persons not easy to identify as such in the presented daylight images. In the meantime the infrared camera is an equal partner to the daylight camera.

First attempts with voice commands detached from the TDS gave an impression on how the workflow could be complemented or simplified. Especially commands which would need several actions with a computer mouse could be done faster by speech, e.g. "show me the image from point X at 12 o'clock". Speech could also enable hands-free entries or requests from the attack teams. During the tests given phrases were recognized independent of the speaker. Focusing the UGVs and their autonomy an improvable point was the reliability of availability. Network instability and some software application problems reduced the use and have to be improved. Latencies in image transmission slowed down the reconnaissance when UGVs were steered manually. Problems with the bandwidth are well known and cannot be solved sure. Therefore the autonomous drive mode became more important. It was performed successfully at the end of the exercise. This emphasizes the benefit of the functionality. UAVs couldn't be used in that area.

TRADR Evaluation (T-Eval) The TRADR evaluation took place in Dortmund, Germany, at the old coal-fired power plant "Knepper" from 26th of October to 03rd of November 2016, organized by the FDDO and followed by an Industry Day on 04th of November at the same place. For the first time the evaluation was for the most part an indoor event. Only during the first run a Falcon from Ascending Technologies, was used for outdoor exploration in the area around the big hall. In contrast to the UAV the UGVs explored the inside of the hall, where GPS signals couldn't be received with the effect, that the UGV localization was not possible. All UGV positions in the TDS didn't match the real position. Without previous knowledge about the location the operator's orientation became challenging. But on the other hand it was a good opportunity for the system and the concept of the system to clarify strengths and weaknesses. The system was evaluated over two days by fire fighters of Dortmund increased by the end-users and partners of the project.

On day one the exercise took place on the ground floor in conditions similar to T-Jex. On a flat plane enriched with obstacles for testing the adaptive flipper control the UGV operators focused on recording for maps as bases for a second sortie in sense of robot collaboration. The operators were also asked to identify points of interest (POI), places with a special challenge as victim or gas/chemical detection. They should evaluate the functionality of the arm in different contexts. Meanwhile the UAV explored the building from outside and send images. Caused by the low smoke density the UAV couldn't detect the sluicing smoke in front of the large gate.

On day two the operation was executed in the big hall on the second floor. UAV and UGVs explored together the area in two sorties. A dynamic element forces a change in the navigation of the UGV between two sorties. It could be demonstrated, that the concept was able to deal with this unknown and challenging area. Due to the mapping and live stream the operator received over the time an impression of the indoor structure and was able to orientate himself. He was able to handle tasks with the robot arm like the grabbing of small packages. But also some weak points were offered. Multiple sensors are necessary to be prepared for different situations. Both the daylight camera and the laser scanner can be disturbed by smoke, but the IR camera can compensate this gap and keep the operator oriented. Similar to the latency for images we could watch a latency for the laser point cloud. Time was needed until the point cloud density was sufficient enough for the operator's awareness. The effect was, that distances respectively the next objective couldn't be estimated well which is especially important for the planning of the autonomous drive.

Despite the technical possibilities the operators didn't use them in the first sortie of day one. This indicates a rise in system complexity which can't be compensated by a 60min introduction. After a more intensive use the operators gain more confidence in working with the system and could apply the autonomous drive and the arm. Due to technical problems with the TRADR Core the functionality of the system was not always given which influenced the operation.

The range of functions raised compared with year 2 and the availability especially of the autonomous drive is an important step towards user-friendliness. Also the mapping and the use of these information for a following sortie is an essential functionality. Both makes the handling of the system significant easier. But with respect to a real scenario the mapping can't be a kind of "pre-sortie" followed by the POI exploration and rescue of victims. This has to go hand in hand.

1.3.6 Summary end-user discussion T-Eval

In this section, the discussions and questionnaires with the firefighters who participated during T-Eval are reported. We asked participants to give feedback on the positive, negative and improvements with respect to the human-robot team for their role. We also asked if they missed any information.

In general, the firemen were positive about the set-up of the scenario; one participant who played the role of TeamLeader had feedback on the assignment he was given. We asked his team to retrieve a sample from the field, but there were also heat sources present. The TeamLeader indicated that he would rather have investigated the heat source with both UGVs instead of grasping and taking a sample. The participants were further enthusiastic to be able to participate in such an interesting evaluation. Most feedback was given by participants on the TDS and the UGV. For a detailed overview of the end-user feedback see Annex [10] (Annex Overview 2.7).

1.4 Relation to the state-of-the-art

To evaluate a large socio-technical system such as TRADR – which is constantly evolving as new functionality is added – in an informative manner is quite challenging. Towards this end we developed an evaluation methodology, see TRADR deliverable DR7.2 [3]. Within robot-assisted search and rescue a number of international competitions exist, the Robocup Rescue League[2], the Darpa Robotics Challenge[3] and the euRathlon robotics competition[4]. The aim of these competitions is to provide common benchmarks against which various functionalities related to robot-assisted search and rescue can be tested. Emphasis in these competitions is a very formal and strict definition of the tasks involved; indeed, to be able to proclaim a winner, strict definition of the rules and judging compliance with these is needed.

Within the TRADR project, typically a choice is made to not partake in such exercises, as this runs the risk of taking away a lot of development time and attention, which can otherwise be spend on our own exercises with end-users, or real life deployments such as Amatrice this year (TRADR deliverable DR8.3). Evaluation within TRADR is more aimed at the holistic performance of the system, rather than measuring more "simple" tasks in isolation. As such we focus on a qualitative description of progress of both various functionalities of the system, as well as progress on running the

[2] `http://www.robocuprescue.org`
[3] `http://www.darpa.mil/program/darpa-robotics-challenge`
[4] `http://www.eurathlon.eu`

integrated system during end-user exercises, details of the latter can be found in Annex [4] (Annex Overview 2.5).

2 Annexes

2.1 Joachim de Greeff and the TRADR consortium (2016), "TRADR Scenario and Use Cases Year 3"

Bibliography Joachim de Greeff and the TRADR consortium, "TRADR Scenario and Use Cases Year 3". Unpublished technical report, Interactive Intelligence, TU Delft, the Netherlands, 2016.

Abstract In this report we describe the scenario and use cases for TRADR Year 3.

Relation to WP This document describes the scenario and uses cases for Year 3. As such it is at the very core of WP7.

Availability Restricted. Not included in the public version of this deliverable.

2.2 Joachim de Greeff (2016), "TRADR Unit UGV Pilot Discussion"

Bibliography Joachim de Greeff, "TRADR Unit UGV Pilot Discussion". Unpublished technical report, Interactive Intelligence, TU Delft, the Netherlands, 2016.

Abstract In this report we describe the discussion regarding potential introduction of UGV pilots.

Relation to WP This document describes the process of rethinking the constitution of the TRADR unit. The outcome of this discussion determined how evaluation exercises were organized and as such contributes to WP7.

Availability Restricted. Not included in the public version of this deliverable.

2.3 Erik Zimmermann (2016), "TRADR Setup"

Bibliography Erik Zimmermann (2016), "TRADR Setup". Page from the internal TRADR Wiki.

Abstract This page describes the setup of the TRADR system and the launch files needed for particular TRADR functionalities.

Relation to WP This page describes in detail the setup of the TRADR system as it was used during T-Eval 2016 and contributes to T7.3.

Availability Restricted. Not included in the public version of this deliverable.

2.4 Daniel Reuter and Mario Gianni (2016), "TRADR Core User Manual"

Bibliography Daniel Reuter and Mario Gianni (2016), "TRADR Core User Manual". Transcript of a live audio recording during the launch process of the TRADR core.

Abstract This report describes the launch process of the TRADR core using the TRADR orchestration tool.

Relation to WP This report is based on a live audio recording during T-Eval 2016 and contributes to T7.3.

Availability Restricted. Not included in the public version of this deliverable.

2.5 Joachim de Greeff and Nanja Smets (2016), "TRADR Scenario-based Evaluation Exercises Year 3: T-JEx and T-Eval"

Bibliography Joachim de Greeff and Nanja Smets, "TRADR Scenario-based Evaluation Exercises Year 3: T-JEx and T-Eval". Unpublished technical report, Interactive Intelligence, TU Delft, the Netherlands, 2016.

Abstract In this report we describe the scenario-based evaluations as executed in Year 3. Specifically, three exercise are reported: T-JEx and T-Eval. We describe the setup and outcomes, along with the end-user comments that were collected during these experiments.

Relation to WP This document reports on the scenario-based evaluation exercises as conduced during Year 3. As such it is at the very core of WP7.

Availability Restricted. Not included in the public version of this deliverable.

2.6 Joachim de Greeff, Nanja Smets, Koen Hindriks, Mark Neerincx and Ivana Kruijff-Korbayová (2017), "Incremental Development of Large-Scale Human-Robot Teamwork in Disaster Response Environments"

Bibliography Joachim de Greeff, Nanja Smets, Koen Hindriks, Mark Neerincx and Ivana Kruijff-Korbayová "Incremental Development of Large-Scale Human-Robot Teamwork in Disaster Response Environments". In Human-Robot Interaction 2017, March 69, 2017, Vienna, Austria.

Abstract We report on the latest large-scale disaster-response exercise conducted by our project, which involves a robotic system with both ground robots (UGVs) and aerial robots (UAVs). In particularly, we focus on aspects related to Human-Robot teaming, and the uptake of new technology by end-users.

Relation to WP This document contains a compact report on T-Eval Year 3, as such it contributes to T7.3.

Availability Public. Included in the public version of this deliverable.

2.7 Nanja Smets (2016), "T-Eval 2016 end-user feedback"

Bibliography Nanja Smets (2016), "T-Eval 2016 end-user feedback". Report on end-user feedback during T-Eval.

Abstract In this report, we describe the feedback we have received from the end-users during T-Eval 2016. The feedback is organized to reflect the positive, negative and improvements mentioned by the end-users on the: TDS, UGV, UAV, TeamLeader and scenario.

Relation to WP This report presents the end-user feedback during T-Eval and is at the heart of WP7 and contributes to T7.3.

Availability Restricted. Not included in the public version of this deliverable.

References

[1] J. de Greeff. TRADR Scenario and Use-Cases Year 3. Technical report, Interactive Intelligence, TU Delft, 2016. Unpublished.

[2] J. de Greeff. TRADR Team UGV Pilot Discussion. Technical report, Interactive Intelligence, TU Delft, 2016. Unpublished.

[3] J. de Greeff, T. Mioch, N. Pahlke, R. Maul, and the TRADR consortium. TRADR Deliverable DR 7.2: Scenario-based evaluation in a large-scale static disaster area. Deliverable, TRADR consortium, 2016.

[4] J. de Greeff and N. Smets. TRADR Scenario-based Evaluation exercises Year 3: T-JEx and T-Eval. Technical report, Interactive Intelligence, TU Delft, 2016. Unpublished.

[5] J. de Greeff, N. Smets, K. Hindriks, M. A. Neerincx, and I. Kruijff-Korbayová. Incremental development of large-scale human-robot teamwork in disaster response environments. In *Proceedings of the Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, HRI'17 Extended Abstracts, 2017.

[6] L. Freda, M. Gianni, F. Pirri, and the TRADR consortium. TRADR Deliverable DR 4.3: Communication and knowledge flow gluing the multi-robot collaborative framework. Deliverable, EU FP7 TRADR / ICT-60963, 2017.

[7] M. A. Neerincx and J. Lindenberg. Situated cognitive engineering for complex task environments. In *Naturalistic Decision Making & Macrocognition*, pages 373–390. Ashgate Publishing Limited, 2008.

[8] N. Pahlke, B. Rohlf, A. Behrendts, N. Smets, T. Mioch, M. Neerincx, J. de Greeff, I. Kruijff-Korbayová, and the TRADR consortium. TRADR Deliverable DR 3.3: Sentient dialogue for tailored SA. Deliverable, EU FP7 TRADR / ICT-60963, 2017.

[9] D. Reuter and M. Gianni. TRADR Core User Manual. Unpublished, 2016.

[10] N. Smets. End-user feedback T-Eval 2016. Unpublished, 2016.

[11] R. Worst and the TRADR consortium. TRADR Deliverable DR 6.3: Multi-robot task adaptation. Deliverable, EU FP7 TRADR / ICT-60963, 2017.

[12] E. Zimmermann. TRADR Setup. Unpublished, 2016.

Incremental Development of Large-Scale Human-Robot Teamwork in Disaster Response Environments

[Extended Abstract]

Joachim de Greeff
Delft University of Technology
Delft, The Netherlands
J.deGreeff@tudelft.nl

Nanja Smets
TNO, Soesterberg
The Netherlands
nanja.smets@tno.nl

Koen Hindriks
Delft University of Technology
Delft, The Netherlands
K.V.Hindriks@tudelft.nl

Mark A. Neerincx
Delft University of Technology
& TNO, Soesterberg,
The Netherlands
mark.neerincx@tno.nl

Ivana Kruijff-Korbayová
Language Technology Lab
DFKI, Germany
ivana.kruijff@dfki.de

ABSTRACT

We report on the latest large-scale disaster-response exercise conducted by our project, which involves a robotic system with both ground robots (UGVs) and aerial robots (UAVs). In particularly, we focus on aspects related to Human-Robot teaming, and the uptake of new technology by end-users.

Keywords

Disaster Response Robots; Persistent Human-Robot Teaming; User-centric Design

1. INTRODUCTION

In recent years, application of robots in disaster environments has increased [3]. However, despite technological advancements, deploying robots in a real-life disaster environment and having them effectively collaborate with humans remains a real challenge. This is partially caused by the fact that it is currently unclear for rescue organizations what can and cannot be expected from rescue robots, due to the rapid development in the field. As such, to ensure effective uptake of these new technologies, extensive tests with end-users remain vital. Towards this end, our project [1] organizes yearly evaluation exercises to align technological developments in the field of Human-Robot teaming with end-users – in our case firefighters – in a structured and incremental fashion. In this paper we report on the latest exercise conducted.

We believe reporting such an exercise is of interest to the HRI community, as it concerns a real-life deployment of a large-scale USAR system under realistic conditions with end users. As such, it provides valuable insights hard to obtain in a more controlled (and therefore less realistic) lab setting.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
HRI '17, March 6–9, 2017, Vienna, Austria.
© 2017 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-4336-7/17/03..
DOI: http://dx.doi.org/10.1145/3029798.3038330

2. REAL-LIFE EXERCISE

The following aspects of deploying robots in a real-life exercise are of particular importance for the project.

Persistence. A real disaster response mission may be prolonged over multiple days, can entail many different sorties (single episode in which the team goes into and out of the hot-zone) and development over time in terms of mission priorities, available information, team structure, hazard levels and environmental factors (e.g. weather etc.). A key aspect for the project is therefore the concept of persistence, i.e. the ability of the system to retain relevant information over time, as to optimally support the rescue effort. In the context of disaster response this includes e.g. deriving and sharing of maps with information of victims, fires and chemical samples, and retain and re-use this information over multiple sorties.

Autonomy. Of particular importance is the level of autonomy that the robots are capable of. The more autonomy they have, the less direct control from human operators is needed, meaning that attention can be spend elsewhere. The UGVs deployed during the exercise were capable of partial autonomous navigation using an automated path-planner. That is, once they were manually driven through an environment and a map was created using SLAM (Simultaneous Localization And Mapping), in subsequent sorties the operator could specify waypoints to which the UGVs could navigation autonomously. Additionally, UGVs could traverse obstacles autonomously through automated flipper control.

Human-Robot teaming. Human-Robot teaming in the context of disaster response operations entails the human rescue workers being able to take advantage of the robot's capabilities when appropriate, while simultaneously not being distracted/hindered by them. Additionally, the use of both UGVs and UAVs increase the systems' complexity. Towards this end, a careful balance needs to be found regarding when to deploy/not to deploy robots, taking into account the aspects of persistence and autonomy as described above.

2.1 Exercise setup

The yearly evaluation exercise took place in a decommissioned power plant in Germany, see Figure 1, left. This environment allowed us to create a realistic disaster response

Figure 1: Left: the indoor environment showing the two UGVs. Right: UGV Operator station.

scenario, which included the following aspects: i) a large-scale, multi-level industrial area to explore; ii) structural collapse, posing a challenge for UGVs to navigate; iii) outdoor and indoor UAV flying options; and iv), simulated fire, smoke, chemical samples and victims. Within this environment a simulated disaster response mission was executed which consisted of multiple sorties over the course of two days.

The rescue team consisted of both human rescue workers, UGVs and UAVs, see Figure 2 for an overview of the team. Human team-members communicated verbally through the TRADR system which was running speech recognition as to provide the system with 'awareness' of what was going on, and through the use of a Tactical Display System. The UGVs exhibited different capabilities, e.g. a mobile arm allowing to manipulate the environment, various sensor arrays and autonomous obstacle traversing capabilities. They were controlled from operator stations located outside of the hot-zone (see Figure 1, right). The exercise consisted of multiple sorties during which various activities related to the mission were executed.

2.2 Exercise outcome

The overall rescue scenario consisted of an industrial accident, in which possible chemical leakage caused an explosion. Possible victims, hazards (fire, smoke, chemical leakage) and structural integrity of the building needed to be investigated. Over the course of two days, 4 different sorties were successfully executed during which the team gradually build up situation awareness of the (simulated) disaster. Using SLAM, the UGVs build maps which allowed them to navigate autonomously, thus easing the cognitive load of the operators. Furthermore, victims, fires and chemical samples were 'retrieved' from the hot-zone, thus contributing to the success of the mission.

On the notion of persistence, the sharing of maps and mission information was well received by the end-users. The aerial overview provided by the UAV helped a lot on the teaming aspects, providing the TeamLeader with a high-level overview, while the UGVs could be used for ground tasks such as communication and sample extraction.

It became also clear that – with an increasingly complex system – sufficient training is a must. End users received about 1 hour training to get acquainted with the system, but appeared that they could benefit from more. As such we will increase training in the next exercise.

Along with the high-level mission objectives, data gathering and evaluation was done for all sub-parts of the system. In particularly, a lot of experience was gained with the creation of maps through multiple sorties, and subsequent use by the automated path-planner, allowing the UGVs to (partially) navigate autonomously. Additionally, we collected valuable feedback from the end-users regarding their ability to operate the system and the extent with which the system enabled them to build up situation awareness, the latter identified as being crucial for effective robot use in USAR environments [2].

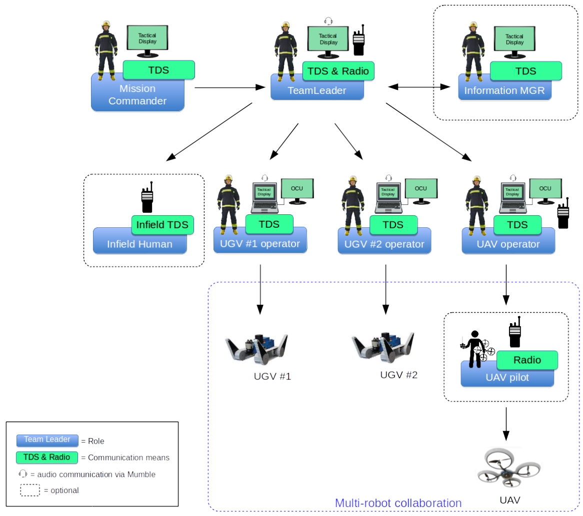

Figure 2: Team structure that was used in the exercise, along with communication and information means.

3. DISCUSSION & CONCLUSION

The execution of a real-life mission allowed us to obtain valuable feedback from the end-users, who are ultimately targeted to operate the system. Generally, the increased autonomous capabilities of the system were received well, with firefighters remarking how such a system can truly be of value in real rescue operations. Also the aerial overview provided by the UAV is generally considered an asset. With the system increasing in complexity every year though, it is becoming more challenging to – when faced with a abundance of data – provide relevant information at the right time to the right actor. This is a non-trivial question which will also be explored in the next year of the project.

4. ACKNOWLEDGMENTS

This work is funded by the EU FP7 TRADR project (grant 60963).

5. REFERENCES

[1] I. Kruijff-Korbayová, F. Colas, M. Gianni, F. Pirri, J. de Greeff, K. Hindriks, M. Neerincx, P. Ögren, T. Svoboda, and R. Worst. TRADR project: Long-term human-robot teaming for robot assisted disaster response. *KI - Künstliche Intelligenz*, pages 1–9, 2015.

[2] B. Larochelle, G.-J. M. Kruijff, N. Smets, T. Mioch, and P. Groenewegen. Establishing human situation awareness using a multi-modal operator control unit in an urban search & rescue human-robot team. In *2011 RO-MAN*, pages 229–234. IEEE, 2011.

[3] R. R. Murphy. *Disaster robotics*. MIT Press, 2014.