Seventh Framework Programme

Information & Communication Technologies

Challenge 2: Cognitive Systems

Grant agreement for: Collaborative Project / Large-scale integrating project (IP)

Annex I – "Description of Work"

Project acronym:	NIFTi
Project full title:	Natural human-robot cooperation in dynamic environments
Grant agreement no.:	247870

Date of preparation of Annex I (latest version): October 29, 2009
Date of approval of Annex I by Commission: December 09, 2009

Contents

Approved by EC on 2009-12-09

List of Tables

List of Figures

Part A

1 Overall budget breakdown for the project

Participant number in this project [11]	Participant short name	Fund. % [12]	Ind. costs [13]	Estimated eligible costs (whole duration of the project)					Total receipts	Requested EC contribution
				RTD / Innovation (A)	Demonstration (B)	Management (C)	Other (D)	Total A+B+C+D		
1	DFKI	75	A	1,692,160.00	0.00	155,680.00	84,522.00	1,932,362.00	159,192.00	1,509,322.00
2	TNO	75	A	1,138,475.00	0.00	48,111.00	66,125.00	1,252,711.00	0.00	968,092.00
3	FRAUNHOFER	75	A	1,247,614.00	0.00	38,685.00	37,885.00	1,324,184.00	0.00	1,012,280.00
4	BLUE	75	A	592,400.00	0.00	46,600.00	48,100.00	687,100.00	0.00	539,000.00
5	ETHZ	75	T	1,162,160.00	0.00	32,800.00	79,680.00	1,274,640.00	0.00	984,100.00
6	CVUT	75	T	803,520.00	0.00	25,040.00	74,880.00	903,440.00	0.00	702,560.00
7	ROMA	75	T	871,233.00	0.00	36,000.00	40,000.00	947,233.00	0.00	729,424.00
8	FDDO	75	T	122,731.00	0.00	7,200.00	0.00	129,931.00	0.00	99,248.00
9	VVFF	50	T	119,520.00	0.00	7,200.00	0.00	126,720.00	0.00	66,960.00
TOTAL				7,749,813.00	0.00	397,316.00	431,192.00	8,578,321.00	159,192.00	6,610,986.00

2 Project summary

NIFTi investigates cognitive architectures which can meaningfully sense, act and cooperate with humans in real-life environments. When it comes to making cognitive architectures "cooperative," research has primarily focused on autonomy, and high-level communication. Little or no attention has been given to making the cognitive architecture adapt to the human – in understanding the environment, planning and acting, communicating. NIFTi picks up on this: *NIFTi puts the human factor into cognitive architectures.*

NIFTi aims to develop *a unified theory of how a cognitive system can achieve natural task-driven cooperation between a human and a robot working together in a dynamic environment, with convincing instantiations on a novel robot platform for urban search & rescue.* The theory rests on the principle of balancing demands on operation and cooperation. This principle guides how the robot needs to act to operate within a dynamic environment, and to cooperate with a human as part of a collaborative activity. Balancing tries to dynamically optimize work flow in the activity, and minimize task load impacts on the human. Striking such a balance in an informed way requires a cognitive architecture to dynamically connect information across a variety of modules. First of all, NIFTi models how a human could operate in a dynamic environment. NIFTi combines models of spatio-temporal structure with observation models to derive a human-oriented conceptual representation of a dynamic environment being explored. This human-oriented representation helps a robot to figure out, how to describe what it sees, and how easy or difficult it may be for the human to execute planned actions (Objective 1, functional environment models). The robot uses these to estimate the possible effects on the human's cognitive task load, and the overall workflow (Objective 2, situated cognitive user models). Accordingly, the robot may decide to adapt how, when and what it communicates to accommodate the human with whom it cooperates (Objective 3, user-adaptive human-robot communication). At the same time, the robot uses the plan and its developing knowledge of the environment to adapt how it operates itself, possibly using new observations to adapt its skills (Objective 4, morphology-adaptive flexible planning and execution). The key scientific breakthrough NIFTi aims to achieve is how a cognitive architecture can facilitate the information flow needed to achieve such a balance, and the insights this brings in formulating (stable) behaviors for operation and cooperation. This addresses directly what is seen as the bottleneck to successful human-robot interaction. A technical breakthrough is the development of a novel rover with adaptive morphology, specifically for the Urban Search and Rescue (USAR) domain.

NIFTi particularly targets end users in the USAR domain, and related domains such as other branches of rescue, safety, and security. The NIFTi consortium consists of several partners with experience in human-robot interaction, human factors and cognitive user modeling, field robotics, spatial and visual modeling of outdoor environments, and flexible planning and execution. The consortium closely collaborates with several end user-organizations. These organizations aid NIFTi in yearly evaluating its integrated systems, providing feedback for further iterations of the development process in NIFTi.

3 List of beneficiaries

Benif.#	Beneficiary name	Benef. short name	Country	Date enter project	Date exit project
1. (crd.)	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH	DFKI	Germany	M1	M48
2.	Netherlands Organization for Applied Scientific Research	TNO	The Netherlands	M1	M48
3.	Fraunhofer Institut Intelligente Analyse- und Informationssysteme	Fraunhofer	Germany	M1	M48
4.	BlueBotics SA	BLUE	Switzerland	M1	M48
5.	Eidgenossische Technische Hochschule Zürich	ETHZ	Switzerland	M1	M48
6.	Czech Technical University Prague	CTU	Czech Republic	M1	M48
7.	‘Sapienza’ University of Roma	ROMA	Italy	M1	M48
8.	Institut für Feuerwehr und Rettungstechnologie FDDo	FDDo	Germany	M1	M48
9.	Corpo Nazionale Vigili del Fuoco	VVFF	Italy	M1	M48

Part B

1 Science & Technology

1.1 Concept and objectives

1.1.1 The aim: Cognition for natural cooperation in dynamic environments

NIFTi aims to develop *a unified theory of how a cognitive system can achieve natural task-driven cooperation between a human and a robot working together in a dynamic environment, with convincing instantiations on a novel robot platform for urban search & rescue.* The theory rests on the principle of balancing demands on operation and cooperation. This principle guides how the robot needs to act to operate within a dynamic environment, and to cooperate with a human as part of a collaborative activity. Balancing tries to dynamically optimize work flow in the activity, and minimize task load impacts on the human. Striking such a balance in an informed way requires a cognitive architecture to dynamically connect information across a variety of modules. First of all, NIFTi models how a human could operate in a dynamic environment. NIFTi combines models of spatio-temporal structure with observation models to derive a human-oriented conceptual representation of a dynamic environment being explored. This human-oriented representation helps a robot to figure out, how to describe what it sees, and how easy or difficult it may be for the human to execute planned actions (Objective 1, functional environment models). The robot uses these to estimate the possible effects on the human's cognitive task load, and the overall workflow (Objective 2, situated cognitive user models). Accordingly, the robot may decide to adapt how, when and what it communicates to accommodate the human with whom it cooperates (Objective 3, user-adaptive human-robot communication). At the same time, the robot uses the plan and its developing knowledge of the environment to adapt how it operates itself, possibly using new observations to adapt its skills (Objective 4, morphology-adaptive flexible planning and execution). The key scientific breakthrough NIFTi aims to achieve is how a cognitive architecture can facilitate the information flow needed to achieve such a balance, and insights this brings in formulating (stable) behaviors for operation and cooperation. This addresses directly what is seen as the bottleneck to successful human-robot interaction [172, 173].

A technical breakthrough is the development of a novel rover with adaptive morphology, specifically for the Urban Search and Rescue (USAR) domain.

1.1.2 The vision: Human-Robot Cooperation for Urban Search & Rescue

Figure 1 illustrates the project vision. To realize this vision, NIFTi adopts a roadmap. The roadmap defines use cases to test specific hypotheses about natural human-robot cooperation. Use cases gradually increase in task- and terrain complexity. (No fire, hazardous materials, etc.) Each use case investigates forms of mixed-initiative cooperation between a human and a ground robot (UGV), in jointly exploring an unknown disaster area. They agree on a joint exploration plan, then keep up a running commentary of what they see and do. The commentary plays a crucial role in managing cooperation, extending and adapting each other's situation awareness, exchanging roles (shared control), and furthering a human's understanding of a robot's capabilities.

The roadmap begins by considering remote cooperation on a single shared task (Years 1 and 2). A human operator sees what the robots see. She coordinates the exploration without physically taking part in it. In these first experiments, NIFTi focuses on human factors that play a role when a human coordinates the execution of a plan, to build up an understanding of the environment shared by human and robot. NIFTi investigates naturalness in such cooperation by examining how a robot's operational demands on autonomous navigation and perception should be balanced against cooperation issues in shared situation awareness, and adapting its autonomy during guided execution of a joint plan. The robot can actively adapt its locomotion morphology to navigate

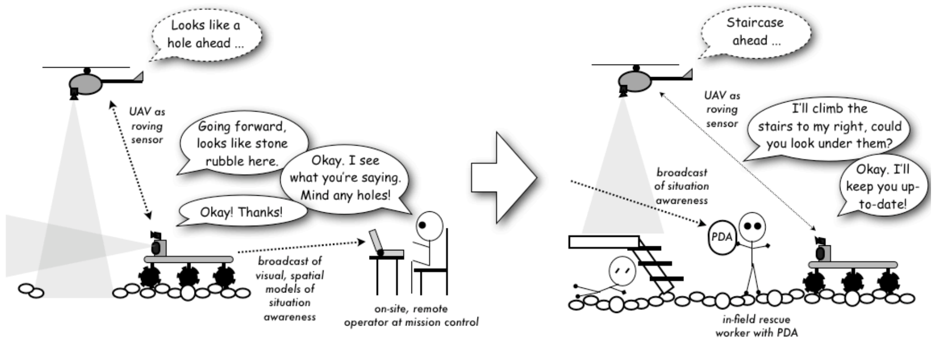


Figure 1: The NIFTi vision: From remotely guided exploration to in-field joint exploration

rough terrain, and use a UAV as roving sensor to actively adapt its perceptual capabilities. In the second half of NIFTi (Years 3 and 4), experiments advance to include in-field joint exploration. Now the human actively takes part in the exploration herself. New issues for the robot arise in understanding how other agents perceive and act elsewhere, adopting different perspectives, and in balancing cooperation demands on action and interaction given that the human herself is involved in her own tasks.

1.1.3 The objectives: Balancing factors in operation and cooperation

The overall agenda of NIFTi is to investigate how cognitive architectures can be designed to include cognitive user models, to make human-robot cooperation natural. NIFTi adopts an operationalization of what it takes naturalness to mean. This operationalization identifies a naturalness loop to achieve a balance between operational and cooperation demands.

The loop is based in the observation that human behavior is dynamic, changing under varying circumstances. In NIFTi, a cognitive architecture learns how and when this tends to happen, so it can adapt the ways it acts and interacts with a human. For this it combines offline, past experience (training, past missions) and online input from the current exploration. The system learns by combining information from the running commentary and perception, with a functional model of the environment. Such a model couples spatiotemporal representation with how such organization makes certain navigation- and perception actions locally possible. Using this understanding of where things are, and how that relates to human behavior during the execution of a joint exploration, the system is able to close the loop. It is able to anticipate how a human may act and interact, so it can adapt how it should cooperate accordingly – making the nature of cooperation fit the circumstances.

The objectives focus on four interrelated aspects of this loop: how humans operate in cooperation, balancing cognitive and situation dynamics; how this effects the form communication takes between a human and a robot, balancing form and content against cognitive load; and, how robots can robustly operate in exploring semi-structured environments, balancing flexible planning & execution, and morphology. NIFTi looks for answers from a human-centered perspective, at the level of integrated adaptive cognitive architectures.

How humans could operate: balancing cognitive and situation dynamics. In NIFTi, robots assist humans in real environments. Cooperation is based on an exploration plan. The human and the robot discuss the plan, and may adapt it as its execution progresses. Creating such a shared plan, and helping to execute it, requires the cognitive architecture to understand what it means for participating agents to be able to perform their parts in the plan.

Objective 1: Functional models of dynamic environments. NIFTi develops new methods for building functional environment models. A functional environment model facilitates a connection between the spatiotemporal organization of the environment, and the conditions for executing an action. This makes it possible to trace and adapt the execution of a plan, given the dynamics and ongoing exploration of the environment. NIFTi combines spatial and perceptual models, with knowledge representations of domain functionality. This includes domain actions (focusing on navigation and perception), objects/landmarks, threats, and how agent morphology can be coupled to domain-specific actions. Formulated a-priori, knowledge can be extended when exploring.

Methodologically, NIFTi achieves Objective 1 by combining insights from robotics (mapping & localization), cognitive science (spatial cognition), and cognitive psychology (situation awareness). NIFTi investigates how a human qualitative sense of acting in dynamic space can be connected to visual- and range data-based modeling of dynamic outdoor areas. NIFTi combines these spatiotemporal perspectives into a single approach to modeling dynamic space, bridging the gap between robot-centric and human-centric concepts. These models are used together with novel methods for connecting planned exploration actions with conditions on the spatiotemporal organization of the situations in which they are to be performed.

A NIFTi cognitive architecture can thus explicitly trace joint plans and adapt their execution as more of the environment is explored. It can connect exploration actions with an understanding of the space in which they are to be executed. In the effort to make cooperation natural, a NIFTi cognitive architecture goes beyond just knowing where and when to act. The system can explicitly reason with how changes in action conditions may positively or adversely affect how a plan is executed. This provides an important bridge to the next objective.

Objective 2: Situated cognitive user models. NIFTi develops new approaches for learning and employing situated cognitive user models. These models connect models of task load and attention with conditions on executing actions in different situations. Using these models in combination with tracing plan execution on functional environment models, the cognitive architecture can anticipate how to act, and adjust focus of attention on what to do next and what to look for.

NIFTi achieves Objective 2 by combining the insights underlying Objective 1 with methodology in cognitive psychology and user modeling. NIFTi investigates how a cognitive architecture can use its growing understanding of the environment, and the corresponding progress of exploration, to predict how a human continues to act. The running commentary between human and robot plays a crucial role here. A cognitive architecture uses the commentary to estimate perceived task load, and to assess what a human is paying attention to in the environment. The system's cognitive user model subsequently combines these two human factors to predict how a human may continue with the plan. Together with Objective 1 this provides a powerful drive towards natural cooperation. Objective 1 provides the NIFTi cognitive architecture with an understanding of why an action and its execution may be (im)possible. Objective 2 turns this understanding into predictions that help the cognitive architecture to anticipate the possible consequences for human action. The NIFTi cognitive architecture can anticipate when a human might adapt the way she acts and interacts. Turning these anticipations into effect by adapting its own behavior, the system can adjust to fit in with a human's dynamic work practice. With this, the NIFTi cognitive architecture goes well beyond just reactively adapting to the current situation.

Methodology in computer-human interaction provides concrete measures to evaluate progress on achieving Objective 2. Anticipation results in less need for reparative measures, such as retracting and replanning failed actions, or clarification of misunderstandings. Specifically, measures for evaluating progress on Objective 2 are task efficiency (less adaptation predicts higher need for clarification, thus longer interactions meaning less efficiency) and the robot's achieved levels of autonomy (less adaptation predicts a higher need for human intervention).

How humans and robots could interact: balancing form, content and situation. Objectives 1 and 2 focus on how the cognitive architecture understands the environment, and how the dynamics of the environment affect how humans act. Objectives 3 and 4 turn this under-

Approved by EC on 2009-12-09

standing into action. When a human and a robot explore an unknown environment, they keep up a running commentary to stay informed about progress. Each agent comments on what she sees and does, dealing with the inherent uncertainty in understanding and acting in a dynamic environment. A key point here is that humans change how they talk, over time. They align how they refer, and what they focus on. This is primarily influenced by how much common ground gets established between the human and the robot, and what effects the ongoing cooperation has on the human's attention and task load. Communication reflects the resulting alignment to load and attention in how broad a scope it takes over the situation, what content is communicated, and in what form this content gets communicated.

Objective 3: User-adaptive human-robot communication. NIFTi develops novel methods for adapting human-robot communication to a user. In NIFTi, such communication is multi-modal (spoken dialogue, GUIs). A cognitive architecture uses its cognitive user models to adjust its strategies for communicating with a user in a given context, to align with perceived changes in communication, cognitive task load and attention. Strategies concern the planning of dialogue (how to continue communication), referential content (what aspects of a plan or the environment to focus on), and realization (in what form(s) to communicate content). Strategies can be parametrized to optimally fit a user, and they can be extended to incorporate communicative routines established during the course of an exploration (referential description, -aliasing).

Objective 3 focuses on the specific challenge of communicating with a human under varying circumstances. The human and the robot operate in an environment that is new and sometimes "defies description." Because the focus is on the task, communication needs to follow in adapting to task dynamics. Methodologically NIFTi achieves Objective 3 by combining insights in robust communication system design from computational linguistics, with insights in the relation between cognitive load and communication coming from linguistics, cognitive psychology, and cognitive user modeling. The resulting designs marry robustness to uncertainty in processing situated communication, with cognitively inspired processes for aligning communication. Alignment focuses on how to balance communication complexity, directing attention, and incorporating routines in understanding and producing communication. A NIFTi cognitive architecture can learn how to flexibly adjust what content is selected and how it is presented, to optimally adjust the complexity of the resulting communication to a user's perceived cognitive task load. Using information from the running commentary, the system can adapt what it pays attention to in perceiving and communicating about the environment. This helps in keeping the focus of attention within communication coherent with what the user is likely to attend to. Finally, the system can extend its competences to incorporate routines that the human and a robot have established so far, notably routines in referring to aspects of a novel environment or a joint exploration plan. Altogether, a NIFTi cognitive architecture learns not only what to say, but also when and how it should say something. How effective this is can be measured objectively, using measures for task efficiency given interaction (e.g. using PARADISE [266]), and subjectively, through user evaluation.

How robots could operate: balancing planning, execution and morphology. Ultimately, success rests on how well a robot can carry out plans. This is where NIFTi closes the loop, between balancing to meet the demands of cooperating with a human, and those arising from autonomously operating in a dynamic environment. It requires a cognitive architecture to be flexible, in adapting its plans and actions to new environments, and adjusting itself to optimally act under different conditions.

Objective 4: Morphology-adaptive flexible planning & execution. NIFTi develops methods for acquiring new strategies for planning and plan execution which explicitly take attention models and morphological constraints into account. Attention models serve to robustly guide planning and execution. They balance a system's perceptual attention between a user-centric perspective on what to pay attention to during exploration, and those aspects of an environment that influence a system's own operation. This balance is achieved by connecting planning and execution with the system's functional environment model and its cognitive user models. Planning explicitly models morphological parameters to take into account what morphological variation is necessary to execute an action in a particular situation. During planning and execution, the cognitive architecture can actively adjust its morphology for locomotion and perception, to meet morphological requirements for executing part of a plan. NIFTi develops a novel robot platform combining active and passive forms of locomotion, and uses a UAV as roving sensor, to investigate the dynamic interplay between a robot's variable morphology and jointly planned exploration.

Objective 4 makes it possible for the cognitive architecture to scale out its capabilities for acting in a complex environment, exploiting the inherent connection between morphology, attention, and flexible planning & execution. While cooperating with a human on a joint exploration plan, the cognitive architecture uses attentional mechanisms to effectively strike a balance between what it needs to focus on to assist a human, and what it needs to watch out for given its own operational demands. It explicitly takes morphology into account. During planning and execution, the system matches action preconditions with the capabilities for navigating and perceiving a morphology provides. A NIFTi cognitive architecture uses this information to anticipate when and how it should adapt its own locomotion or perceptual morphology, or to dynamically react to changing or unforeseen environment conditions.

1.1.4 The design: From the integration of ideas to real-life scenarios

NIFTi investigates how cognitive architectures can cooperate with humans in a *natural* way. NIFTi adopts the standpoint that this can only be achieved if we take into account the human dimension: *The human factor is the measure for meaningfulness in human-robot cooperation.* NIFTi operationalizes the idea of naturalness to make clear the balance a cognitive architecture needs to achieve between what it needs to do itself, and how it can meet the demands of cooperating with a human. This balance is *dynamic*, a dynamics NIFTi specifies as the Naturalness Loop. The objectives address the different aspects of this loop.

Table 1 describes how the NIFTi objectives contribute solutions to several research questions for Objective 2.1. NIFTi achieves these solutions through consistent *integration of ideas across the project first, and then instantiating and evaluating methods and systems in real-life scenarios.*

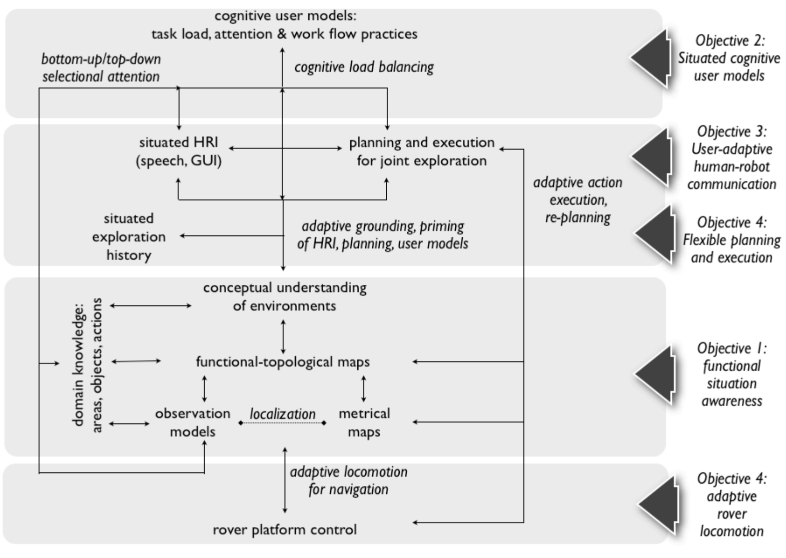


Figure 2: The NIFTi cognitive architecture design

NIFTi develops an integrated cognitive architecture as shown in Figure 2. The design captures

Question	Contributing to solutions
High-level cognitive skills	**Beyond passing information**: A NIFTi cognitive architecture locally controls the dynamics of information flow and attention between processes. Controllers actively decide how to interconnect information across modules, to ensure that changes in observation and interpretation are dynamically percolated bottom-up and top-down to achieve coherence across different levels of understanding, anticipation, and behavior. **Concerted user adaptation**: A NIFTi cognitive architecture uses conceptual representations in the functional environment model as common ground for exchanging and interconnecting information between communication, flexible planning & execution and cognitive user models. Interconnection guides attentional processes and adaptation to operational and cooperation demands.
Autonomy	**Cooperative autonomy**: A NIFTi cognitive architecture adapts when and how it behaves autonomously by balancing operation- and cooperation demands, to optimally fit with work practice (joint plan, scheduling actions) and cognitive task load. **Attentive autonomy**: A NIFTi cognitive architecture adapts what to explore, to balance its own information demands (arising from operation given a plan) and the anticipated or agreed-upon information demands of a human agent.
Interaction modeling & design	**Adaptive alignment**: A NIFTi cognitive architecture adapts the scope, content, and form of communication dynamically, to align with the cognitive task load of a user and the common ground (established referring descriptions, aliases).
Learning	**Interaction-driven learning**: A NIFTi cognitive architecture uses offline past experience together with current interaction with a user (running commentary, joint exploration) to adapt its cognitive user model, and by coupling a cognitive user model with planning, execution and communication it learns how to optimally balance demands **Exploration-driven learning**: A NIFTi cognitive architecture uses its experience in navigating and perceiving dynamic environments to adapt and extend the perception and navigation skills in its action planning domain.
Natural language	**Context-sensitive comprehension**: A NIFTi cognitive architecture uses contextual information (perception, planning, communication, active domain knowledge) to prime the processing of spoken dialogue, and guide the interpretation of incomplete or ungrammatical utterances in context. **Context-sensitive production**: When aligning to a user, a NIFTi cognitive architecture uses contextual information (small- and large-scale spatial context, planning, domain knowledge, dialogue context & common ground) and linguistic models of cognitive processing load to appropriately determine how to describe aspects of the environment.
Representation and modeling	**Functional understanding of environments**: A NIFTi cognitive architecture actively combines domain knowledge about actions and their preconditions, with plans, morphological capabilities, and hybrid representations of the (observed or expected) spatio-temporal structure of the environment to build up an understanding of where things are, and how and when the environment makes it possible for a particular robot / morphology to execute a specific action. Together with cognitive user models this forms the basis for a genuine Endsley-like notion of situation awareness [50, 51].
Rich sensory-motor skills	**Cognitive execution**: A NIFTi cognitive architecture actively combines attention, planning, and functional environment models to acquire skills for cognitive execution – scheduling and balancing operations (perception, locomotion). These skills guide adaptation, task switching, attention and mixed-initiative from an operational view.

Table 1: Objectives: How integration of ideas contributes to pertinent research questions

the closely coupled interaction between the environment models a system builds up, and how these models provide a basis for interconnecting conceptual representations across cognitive user modeling, communication, planning, and a situated exploration history. The latter provides a spatiotemporally grounded event memory of what was seen, said, and done by the agents involved in a joint exploration. Changes in perception and action can be percolated throughout the architecture, with information remaining grounded and interconnected across different modalities. This results in adaptation being both local, and system-wide. Locally, controllers use attention mechanisms, modality-specific offline and online learning algorithms, to guide information flow between modalities. At the same time, controllers ensure that information remains interconnected and situationally grounded across modalities. Adaptation to change, whether triggered externally or internally, thus ultimately emerges as a coherent system-level phenomenon.

NIFTi adopts various methodologies to achieve its objectives, contributing to the design in Figure 2. A core role is played by the functional environment models. Together with the cognitive user models, they form the basis for the architecture's situation awareness. The functional environment models are a collection of interconnected perceptual and spatiotemporal models. The controllers linking these models are bidirectional, so that information can flow bottom-up and top-down. This makes it possible for perceptual information to percolate upwards and trigger domain inferences – and, going top-down, to use domain inferences and mission briefs about expected spatial aspects to prime selectional attention mechanisms in active vision.

The NIFTi cognitive architecture has access to various perceptual modalities (omni- and stereo-cameras, 3D-laser) to construct these models. The models use an autonomously acquired metrical map of the environment (2D, evolving to 3D) as base reference model. The metrical model is annotated with instances of recognized terrain features, static landmarks, dynamic events, and threats. The NIFTi cognitive architecture uses the running commentary, as well as associations between perception, actions, and domain knowledge, to create qualitative representations of space. The first level of qualitative abstraction over the metrical map is a topological map. The topological map captures topological organization over regions (e.g. left, right, before, over, under). The NIFTi cognitive architecture combines metrical map information (terrain features, landmarks, events, threats), topological organization, and inferences over domain ontologies to build up conceptual representations of areas. Domain ontologies link regions with spatial aspects to be expected there (e.g. landmarks or threats), and spatial aspects to actions. The NIFTi cognitive architecture combines the inferred possible actions with the exploration plan – the "circumstances"' under which the plan is to be executed. The system uses these circumstances to check whether and how an agent can execute them, and to estimate the associated cognitive task load and attentional focus. The combination of conceptual and topological information provides the basis for grounding several types of information. For example, the system relates topological regions with information in the situated exploration history and perceptual information ("this path leads to an observed threat"), to plans ("this path is a prioritized choice in exploration"), and to domain inferences and dialogue ("these regions may contain victims"). The domain ontologies are pre-defined, not learnt from scratch. NIFTi uses standardized ontologies for USAR and similar domains. The focus in NIFTi is on how the cognitive architecture can use these ontologies, learning how they can be applied to spatial organization (Objective 1) and to what extent actions inferred as possible can indeed be executed (Objectives 1, 4). The ontologies are human-defined ontologies. Enabling the cognitive architecture to ground them in its own experience, these ontologies help bridging the gap between robot-centric and human-centric perspectives.

The NIFTi cognitive architecture integrates core functionalities common to collaborative flexible planning, and dialogue. They share how their content is connected to the environment models, to cognitive user models, and to the situated exploration history. Again, information flow is bidirectional. Planning and dialogue contexts can prime what to look for in the environment. Vice versa, what is known about the environment primes what to do, how to understand what a human says, and how the system can describe or refer to aspects of the environment. The bidirectional connections between these different levels of understanding use a common layer of abstraction which includes adaptive methods for reference resolution. These methods can resolve references to spatial organization, individual spatial aspects, and spatiotemporal events. Resolution is against the environment models and the situated exploration history. Broadly speaking, context including user models determines how plans and dialogues are understood and produced.

Using its knowledge about the environment, and how the joint exploration plan is progressing, the NIFTi cognitive architecture estimates the user's cognitive task load and attention. The architecture uses these estimates to determine how and when it should say or do something (Objectives 2–4). By keeping up a running commentary, the cognitive architecture and the human know what each is doing. If at some point the architecture cannot proceed, is it better to immediately request assistance, or wait until the human is done? And, if it needs to refer to a spatial aspect, how best to do that so it is easy for the human to understand what the system means? When should it use short utterances, with relatively simple content – and when should it be more elaborate? Estimating cognitive task load in a given situation, against the background of an exploration plans, helps the cognitive architecture to determine how best to behave.

When the cognitive architecture executes actions to explore the environment, all aspects of the architecture interact. The architecture continuously monitors plan execution, comparing the achievability of planned actions against what is now known about the environment. At the sensory-motor levels, the architecture uses its understanding of the environment to anticipate necessary changes in its means of locomotion or perception, and to avoid obstacles while navigating (Objective 4). As the environment models get updated during exploration, the NIFTi cognitive architecture keeps up its side of the running commentary to keep the human informed.

1.1.5 The roadmap: Evaluation on real-life scenarios

The NIFTi project partners have established world-class reputations in their respective fields, and the combined expertise covers the necessary basis for addressing the project objectives. Table 4 (p.108) presents for each partner the contributed expertise and research; (cf. §2.2-§2.3). Several partners have experience with the USAR domain, having won awards at the RoboCup Rescue competitions, and having worked with USAR end user organizations – for example at the March 2009 disaster in Cologne (collapse of the city archives building). This prior experience provided the basis for the NIFTi S&T roadmap. NIFTi instantiates its scientific theories as fully integrated systems. The theories drive the progress, and it is the real systems that measure that progress. The roadmap captures this interplay between the project's scientific and technical objectives through the integrated systems, *formulated as yearly milestones*. Each milestone focuses on questions which arise from NIFTi's operationalization of natural human-robot cooperation. NIFTi involves multiple end-user organizations to help in performing domain analyses, formulating use cases, and doing the yearly integrated system evaluations. The USAR domain provides NIFTi with realistic settings to focus the scientific problems, evaluations help to focus further progress.

The roadmap follows a single scenario concept. Rescuers need to make an assessment of the real situation at a disaster site. NIFTi systems are deployed (UGV, with a UAV as roving sensor) to explore quadrants of the disaster area. A rescuer teams up with a UGV. The robot builds up its environment models, and communicates with the rescuer. Relevant information is displayed on the rescuer's GUI (a PDA in-field, or a remote laptop), e.g. topological and conceptual mapping information, visual information about landmarks, terrain, or threats. At any time, the rescuer can use speech or the GUI to intervene in the robot's actions, or discuss where to explore further.

To achieve robustness in cognitive system functionality, the roadmap defines use cases to introduce the necessary variation into the scenario concept. The use cases define gradually more complex missions which test NIFTi's working hypotheses. Firstly, this results in *variation in the mission* which a robot and a human are to jointly carry out. NIFTi starts with human-instructed exploration, in which a remote operator guides a robot through multi-modal interaction (Yr1), followed by human-assisted exploration, in which the robot is more autonomous and jointly plans exploration (Yr2). NIFTi subsequently increases the robot's initiative, so it can share its situation awareness with other agents in the field – first to plan joint exploration with a rescuer in-field (Yr3), and then to jointly explore and share situation awarenes (Yr4). Secondly, the use cases set these missions in gradually more complex terrain, thus achieving *variation in terrain*. NIFTi deploys and evaluates its systems in real-life training areas of complexity comparable to NIST USAR arena types Yellow (Yr1), Orange (Yr2), and Red (Yr3,Yr4). Thirdly, NIFTi involves several end-user organizations to achieve a *variation in users*.

Roadmap Yr1: human-instructed exploration. Cooperation involves a human instructing a robot how to explore an environment of NIST USAR Yellow level. The robot autonomously navigates, executing the exploration plan, and communicating what it sees. The **human factor** focuses on *user modeling*. Central questions are, How does a human's cognitive task load vary, when building up situation awareness based on the robot's experience? How does the use of spoken dialogue improve task load during exploration? What does a human pay attention to? To assess and predict human instruction performance, NIFTi models the human's focus of attention and task load, and the chances for performance deficits. The models capture the complexity of the environment for the human, and the effects of robot behavior, environmental conditions and events on human performance (WP4). To add time-pressure, the robot needs to completely navigate the area in under 10 minutes. To **operate**, the robot semi-autonomously acquires layered 2.5D spatial models, capturing spatial organization, (visual) landmarks, and terrain classification for traversability and threat analysis (WPs 1,2). Communication focuses on describing space, and instructing a robot where to go (WP3). The robot can adapt navigation and perception actions (WP5), and ground human descriptions of static spatial aspects (WPs 3,5). HRI in year 1 is on-site, but remote. NIFTi develops a single integrated cognitive robot architecture (WP7). NIFTi **evaluates** what a robot needs to understand about its environment, to help a human assess a disaster situation, and how to communicate that understanding (WP3–5); and, suitable morphologies for moving in semi-structured areas (WP6). Performance measures are the actual time spent in the area, task efficiency, and map quality (organization, landmarks, threats, terrain features, and victims found) (WPs 4,7).

Roadmap Yr2: human-assisted exploration. Cooperation extends to mixed-initiative interaction for human-assisted exploration. Environment complexity increases to NIST USAR Orange level with dynamic threats. The robot navigates autonomously, using attention to drive exploration. The **human factor** focuses on *user adaptation*. When and how should the robot adapt autonomy, given human cognitive task load in building up situation awareness? How should communication, attention be adapted to the human? The general hypothesis tested this year is that better adaptation yields better overall performance of a human-robot team. To **operate**, the robot's autonomy in exploring and interacting is increased. NIFTi extends spatial models to capture 3D structure (WPs 1,2), acquired using attention-driven autonomous exploration (WPs 4,5). Multi-modal HRI focuses on jointly establishing exploration plans for the robot, with the (remote) operator possibly intervening or instructing the robot while exploring (WPs 3,5). HRI extends to include adaptive grounding to spatiotemporal descriptions (WP3), and task load-driven adaptation of dialogue style (WPs 3,4). Planning adapts actions based on navigating more complex terrain (WP5). NIFTi investigates linking variable morphology with situation awareness (WPs 1,2,5,6). The integrated cognitive robot (WP7) has variable morphology, mixed-initiative dialogue, and a degree of autonomy sufficient to explore the environment with only communicated assistance. **Evaluation** further investigates proper forms of communication (task efficiency), situation awareness (map quality), adaptive locomotion morphology (complete navigation, given unrestricted time), and task load (user's situation assessment and robot control).

Roadmap Yr3: in-field joint exploration planning. Cooperation becomes in-field. An in-field rescuer communicates with a robot to establish a joint exploration plan. The **human factor** studies the effects of *cooperation* on building up situation awareness. The hypothesis is that the running commentary facilitates human-robot cooperation in field, lowering cognitive task load while simultaneously improving the human's situation awareness. Particular focus is on dealing with perceiving and describing the environment from different perspectives, adaptation to cognitive task load, and coordinating ongoing (individual) exploration tasks. For **operation**, the cognitive architecture is extended to handle understanding the environment from different perspectives. This includes visuo-spatial understanding (WPs 1,2) and understanding a human's attention (WPs 3–5). HRI and planning are able to adaptively understand, and produce, perspectivized descriptions when establishing a joint exploration plan (WPs 3,5), providing feedback on task load dynamics (WP4). Adaptive, multi-modal HRI extends the work on jointly establishing exploration plans for the robot (WPs 3,5) in the context of USAR field practices (WP 4). Variable locomotion and perception morphology is extended to deal with environments of NIST USAR Red complexity. The integrated system includes these functionalities (WP7), and is tested on complete navigation of red-level areas without time limits.

Roadmap Yr4: sharing situation awareness. **Cooperation** extends interaction with in-field operators, investigating how an operator can cooperate with a robot to build up a shared situation awareness. The **human factor** focuses on active cooperation, investigating *working agreements* to establish how a shared awareness can be efficiently build up during joint exploration. Working agreements further robot behavior and communication adaptivity, to fit in with human practice. Environment complexity remains at NIST USAR Red level. For **operation**, the cognitive architecture acquires and maintains situation awareness using the robot's own experience, and communicated experience (WPs 1–3). The robot adapts to, and integrates, situational information of differing spatiotemporal referential nature, at different levels of detail, trying to understand how the environment is seen from different perspectives while two agents are acting and paying attention to specific aspects of the environment (WPs 1–5). The integrated system includes these functionalities (WP7). NIFTi provides a GUI to set-up robot's adaptive behaviour and a function for collaborative mapping that aligns with a human's attentional focus and task load (WPs 3,4). **Evaluation** is extended to include how a robot can comprehend and produce characterizations of situational awareness, and use shared situation awareness to guide its own actions (WPs 3–7). Task-dependent time limits are imposed for navigating the area.

1.2 Progress beyond State-of-the-Art

Contributions to progress. "Whereas early research on teamwork focused mainly on interaction within groups of autonomous agents or robots, there is a growing interest in better accounting for the human dimension. Unlike autonomous systems designed primarily to take humans out of the loop, the future lies in supporting people, agents, and robots working together in teams in close and continuous human-robot interaction." (Sierhuis & Bradshaw, p.c. 2009)

NIFTi focuses on how such continuous human-robot interaction can be made natural, in a cooperative setting. NIFTi operationalizes "naturalness" as finding an optimal, dynamic balance between the operational demands of a robot, and the demands arising from cooperating with a human. With this approach, NIFTi contributes to the state-of-the-art in cognitive systems in the areas of situation awareness (including the combination of robot mapping, spatial cognition, and cognitive user modeling), human-robot interaction, flexible planning & execution for cooperation, and adapting robot morphology. Table 2 (p.21) summarizes the progress NIFTi intends to make (with references to state-of-the-art discussions, and work packages where indicated progress is achieved).

How humans (and robots) could operate: Functional environment models. If a cognitive system is to understand how it could cooperate with humans, it should first of all understand how a dynamic environment might influence its own operation and that of a human. This is what functional environment models aim to do. Taking the perspective that "mapping is for action," NIFTi constructs models of that environment that explicitly connect *what* is *where*, to *how* such organization might influence action. Thereby, NIFTi focuses particularly on navigation- and perception actions, bearing in mind the roadmap. §1.2.1–§1.2.3 place these ideas in the current state-of-the-art, and outline the intended contributions. Together, these contribute to achieving Objective 1. §1.2.2 describes spatiotemporal models. These models connect $3D^+$-metrical representations to qualitative models of topological structure, and human-oriented conceptual information about such a dynamic structure. §1.2.3 describes how this structure is further populated with information about perceivable objects, events, and threats. Functional information is then established at the interface between spatiotemporal models and observation models. Conceptual inference combines information about observed occurrences (what, where) to infer effects on navigation- and perception actions, and then projects these effects back into the spatiotemporal models (how) using occurrence information (pose, position).

How a human could operate: Integrating cognitive user models. A cognitive user model describes factors that influence how well a human can perform a particular action in some situation. This information is key if a robot is to cooperate in a natural way. Ideally, the robot-as-assistant would act such that the human can optimally perform when cooperating with the robot. §1.2.5 describes cognitive user models in more detail, focusing on cognitive task load and attention. NIFTi situates these models by connecting them to the functional environment models, and the action plan that the human and the robot have established (§1.2.7, discussed below). For this, NIFTi uses a cognitive architecture that is described in §1.2.4. This cognitive architecture provides the means to combine information across different levels of representation in a dynamic fashion, and the (probabilistic) mechanisms for learning how and when such connections can be established. Taken together, this contributes to achieving Objective 2.

How a human and a robot could interact. In NIFTi, a robot and a human communicate over an open-loop using spoken dialogue, and a multi-modal GUI. Typically, they talk about the environment, what is where, what is happening, and what they are doing. Depending on the situation and what each is doing, this communication can take different shapes: how much content is conveyed, what is focused on, when something is said. The robot's cognitive system can use its situated cognitive user models and estimated user state to guide how and when it should adapt the ways it communicates with the human during cooperation. This type of user adaptation, and the way communication gets dynamically connected with the different models a robot maintains, are contributions made by NIFTi. §1.2.6 places these issues in the context of the state-of-the-art in human-robot interaction, focusing on dialogue and symbol grounding. The latter point expands on the earlier discussion on symbol grounding (§1.2.4). In dialogue, grounding needs to take referential content into account (resolving what an expression refers to), and communicative routines (e.g. referential aliasing). Addressing these issues together contributes to achieving Objective 3.

How a robot could operate. Finally, the previous themes are connected to operation. §1.2.7 discusses cognitive models for planning, executing, and acquiring actions. These issues cut across the robot's own operation, and how it co-operates together with the human (i.e. other agents). A novelty in NIFTi is that a robot can use developmental forms of learning to acquire new skills, for performing actions and for controlling their execution in a cooperative context. This provides one means for adapting its operation. Another means regards adapting the very morphology the robot uses to navigate and perceive in a particular situation. §1.2.8 discusses adaptivity at the level of robot control and locomotion. NIFTi presents a new robot platform with both passive and active forms of locomotion, which will be light enough to be carried by one or two persons. The exact weight (estimated at 25kg or less) needs to balance the typical low weight of space rovers, and size/weight requirements for operating in rough terrain. The robot has an UAV to be used as roving sensor. Together, addressing these issues in adaptive control, planning and execution contribute to achieving Objective 4.

Technically, there is an explicitly identified need for the robots which NIFTi envisions. Day-to-day work of rescue workers like fire fighters includes the localization of victims as well as fires and other hazardous materials in homes, factories, warehouses, and other buildings, such as hotels, swimming pools and sports halls. Search and rescue is a dangerous and people-intensive task. Self-security devices and personal protective clothing of fire fighters leads to an encumberance of approx. 30 to 40 kg and to a restricted perception. This task is further complicated by the presence of adverse conditions, such as working at night with little or no electrical light available, or smoke emissions that hinder the sight. In order to correctly assess the situation, fire fighters have to rely on special sensors, such as measuring devices for gases or radiation, as well as thermic image cameras. This leads to even more weight being carried and might cause a tunnel view effect where the fire fighter loses his situation awareness due to a cognitive overload. Robots have been found to provide helpful assistance in search and rescue. Using suitable sensors, the robot can itself attain a degree of situation awareness that allows it to take initiative within a sliding autonomy framework. Using the right methods, the information gathered by the robot can be presented to its operator in a way that increases the operator's situation awareness while keeping (additional) cognitive workload minimal.

The robots for USAR NIFTi envisions present more than just a technical solution. Sociological, demographical, economical and technical factors influence the problems fire departments will be dealing with in the near future. Fire brigades are facing more and more the problem of limited availability of man-power: less people join voluntary fire brigades; professional fire departments suffer from the fact that public authorities have less money to spend on man-power. Moreover, the demographical structure of western societies will change over the next centuries. The quintessence of this is that in an ageing population, less people will be working to sustain an adequate living standard. For rescue workers this means that there will be even less active people available and those will have to be kept away from life threatening risks even more. On the technical side, the materials used to construct houses, machines and vehicles, and the substances used in factories and plants will lead to more dangerous situations in case of an emergency, e.g. more smoke, hotter fires, more rigorous structures that are harder to break through etc. A major challenge for fire brigades consists of damage containment. The source of a fire or other damaging event needs to be located as quickly as possible in order to minimize collateral damage to buildings, machines, and most importantly human beings. A robot that actively locates victims can help safe more lives, e.g. through leading rescue workers to victims in a fast and safe way.

The possible contributions of robots are thus not only in augment situation awareness of rescue workers. If robots act as semi-autonomous mobile sensors and mapping systems, rescue workers and fire fighters can be assigned more efficiently to other tasks. Thus more can be achieved by the same number of people. A robot that actively locates hazardous environments can help prevent rescue workers from being harmed in the line of duty, e.g. the robot can be sent to explore areas that are too dangerous or inaccessible for human rescue workers. A robot that actively locates the cause of damage can help limit collateral damage, e.g. through minimizing the damage that aerosols cause on machines that are not directly affected by the damaging event.

The need to go beyond the state-of-the-art. The NIFTi objectives are based in current research, but ultimately require NIFTi to move beyond the state-of-the-art. Past years have seen significant progress on the several issues pointed out above, in diverse fields as AI, language processing, cognitive science, computer-human interaction, computer vision, and robotics. The project partners often significantly contributed to these developments. Current integrated robot systems that use these advances still have one or more restrictions on their abilities, though:

- Environment models are built up in a purely bottom-up fashion, and usually rely on clearly defined structure in an environment (e.g. walls, doors of specific sizes).
- There is only a limited integration of environment models (maps; vision) and dialogue, planning. If at all, integration relies on simple, clearly defined mappings, and there is usually no top-down attentional priming of attention of perception on the basis of dialogue- and task-context.
- Task-setting is driven mostly by the human – either in a very strict sense, through tele-operation, or through very restricted forms of spoken dialogue.
- Cognitive architectures hardly adapt their system behavior (acting and interaction) to the user.

Although current research is slowly overcoming these limitations, there is no unified solution that would enable NIFTi to achieve its project objectives using current state-of-the-art. NIFTi needs to move beyond the state-of-the-art in a number of areas. Particularly, NIFTi needs:

- Cognitive architectures which integrate cognitive user models as a fundamental factor affecting understanding and behavior, (not as an isolated module).
- Cognitive architectures that are based on mechanisms for connecting content across different modalities, with the explicit purposes to (a) disambiguate and complete information, (b) prime attention mechanisms in one modality, on the basis of salient information in other modalities, and (c) prime deliberate decisions for acting and interacting based on user models including task load and attention.
- Representations and algorithms for relating quantitative and qualitative information, possibly across multiple layers of abstraction, in specific domains of spatial models, observation models, (situated) dialogue, flexible planning, and decision processes. Making these representations adaptive, to adapt to changes in the environment and novel ways for reference and description.
- Cognitive architectures with the ability to adapt low-level behaviors for perception and action on the basis of high-level contexts (dialogue, planning), to allow for a mixture of reactive and anticipatory behaviors.
- Robots with the ability to autonomously carry out partially, qualitatively specified tasks by planning mobile movement, selective information acquisition for situation awareness, and information processing in a unified cognitive architecture.
- Cognitive architectures with the ability to communicate with a human to jointly resolve issues in uncertainty or incompleteness in understanding the environment, or action outcomes.
- Cognitive architectures with the ability to assess a situation not only from its own, ego-centric viewpoint, but also from other perspectives, to be able to cooperate in a shared environment.

The next sections describe the relevant current state-of-the-art, and how NIFTi moves beyond. Table 2 summarizes NIFTi's main advances.

1.2.1 Situation awareness

Situation awareness (SA) represents an understanding of what is going on in an environment, as far as relevant to a particular goal or task [50]. Particularly, for robots NIFTi considers SA to be "the *perception* of the elements in the environment with a volume of time and space, the *comprehension* of their meaning, and the *projection*" of their future state [49] (Figure 3). This entails three levels of SA (perception, comprehension, predictive projection) which altogether form the basis for decision making, human-robot interaction, and user adaptation [46].

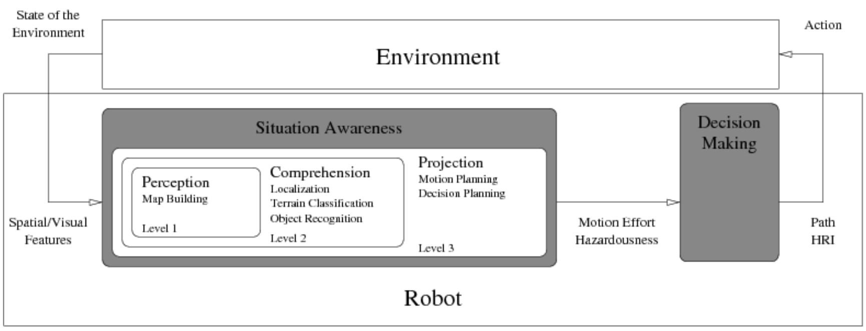


Figure 3: Model of situation awareness applied to robotic domains.

NIFTi focuses on several types of models for understanding the environment: Models of the spatiotemporal structure of the environment (the *where* and *when*; §1.2.2), and observation models that represent sensory perception of objects and events (the *what*; §1.2.3). These two aspects

Objective	State-of-the-Art	Long-term goal	State-of-the-Art after project
Objective 1, functional environment models	Models of static spatial organization	Affordance-based spatiotemporal understanding of dynamic indoor- and outdoor environments	Projection of domain-specific actions and their anticipated effects into hybrid spatiotemporal models of disaster environments
	SotA: §1.2.1–§1.2.3		WP1 (§1.3.5.1), WP2 (§1.3.5.2)
Objective 2, situated cognitive user models	Autonomy, communication, planning and attention are insensitive to human task load, attention, work flow	Cognitive user models are grounded in situation awareness, collaboration, to drive aligning with a human's perceived and inferred task load, attention work flow	Cognitive user models drive adaptation in communication, planning to align with human-commented and inferred task load, attention, work flow in USAR.
	SotA: §1.2.5		WP4 (§1.3.5.4))
Objective 3, User-adaptive human-robot communication	Controlled "tightly scripted" HRI for acting in current scenes, navigation of static environments.	Free HRI for team-based collaboration in dynamic indoor- and outdoor environments	User-adaptive HRI for human-robot joint exploration in dynamic disaster environments
	SotA: §1.2.6		WP3 (§1.3.5.3), WP4 (§1.3.5.4)
Objective 4, morphology-adaptive flexible planning & execution	Model-based executive control: flexible concurrent plans are engaged, disengaged in a reactive control loop. Combining planning, scheduling and resource-optimization to manage competing activities.	Handling response stimuli for behavior interaction, temporal switching between events, task-driven attentive execution. Coherent models for concurrent, context-sensitive flexible plans.	Resource allocation for sensorimotor- and task-selection processes. Mapping between internal states and execution. Flexible behaviors, morphology adaptation of locomotion, perception.
	SotA: §1.2.7, §1.2.8		WP5 (§1.3.5.5), WP6 (§1.3.5.6)

Table 2: Project objectives and progress on state-of-the-art, including references to sections discussing state-of-the-art and progress (SotA), and the WPs in which indicated progress is achieved.

are interrelated: Categorical understanding of objects and events reveals more about their functionality, and potential effects. NIFTi projects this information about functionality and effects into spatiotemporal organization, to understand how the environment enables agents to act, with anticipated task load and attention (the *how*). Together, §1.2.2 and §1.2.3 regard the efforts aimed at achieving Objective 1.

1.2.2 Spatio-temporal models for situation awareness

Humans and robots have inherently different representations of space. A robot map needs to accommodate safe and reliable motion control, to preserve a good self-localization, and to provide a basis for path planning. These specialized, *quantitative*, spatial representations of the robot are ill-suited for intuitive communication with human users. It has been shown that people rather use *qualitative*, e.g. topological , spatial knowledge when reasoning and communicating about their environment. For natural human-robot interaction this gap must be bridged by constructing *multi-layered maps* which combine quantitative and qualitative, spatial and conceptual information of different granularity at different levels of abstraction.

Learning a spatial representation of the environment is an important issue in robotics in which maps are build mostly with 2D range scanners. With such sensors, the representation can consist directly on raw data points or meshes, which allow for hybrid maps with the surimposition of visual information. However, the memory requirements are high especially when mapping a large environment with a sensor on board of a vehicle. Techniques such as piecewise linear approximation, 2D or 3D grids, or tree-based maps are then used to build a spatial environment representation.

Capturing world dynamics in *spatio-temporal environment representations* ranges from multi-target change detection to multitarget tracking problems. Data association between subsequent observations presents the central problem. Most data association methods have an exponential complexity that can be overcome by approximation strategies.

Closely related to mapping is the issue of *robot self-localization*. This problem is usually studied in the context of *simultaneous localization and mapping* (SLAM). Most techniques extract landmarks from range or vision data, calculate the map and the pose of the vehicles based on these landmarks, estimating a robot's pose in all six degrees of freedom.

Conceptual information is crucial if robots are designed to have a situation awareness that goes beyond self-localization and simple path planning. Human-built environments tend to cluster space into regions that afford related functionalities, and this is also reflected in the common names for areas (e.g., a kitchen is a place where one can prepare food, and a parking lot is a place where cars are parked). Knowing what a region is called usually allows humans to understand what they can do there and which objects they can expect to find there (such as, e.g., a sink, an oven, or pans in a kitchen) – and vice versa: perceiving what actions are afforded in a place allows humans to categorize and name it. An appropriate spatial representation is a necessary precondition for a successful and efficient human-robot communication. In order to provide a common ground for interaction with a human, such a spatial representation must contain spatial units that resemble the way in which humans segment space, and it must allow for referring to those spatial units in a way that is natural for humans. In NIFTi we will develop maps that are based on a functional spatio-temporal understanding of space. This new kind of map will allow the robot to interact adequately with both humans and the environment, and provide necessary steps towards a fully affordance-based approach for spatio-temporal understanding of dynamic environments. Function-based clustering of space will enable the robot to give structure to otherwise unstructured environments, such as a disaster area. This provides a good basis for efficiently conveying situation-awareness to remote operators, as well as for setting appropriate frames of reference when instructing the robot. Since this resembles the way humans conceptualize space, it allows us to keep the operators' cognitive taskload for communicating with the robot low. According to the classification in [30], a conceptual hierarchy grounded in the spatial hierarchy is assumed to contain semantic and conceptual information about space at different levels of abstraction.

Topological maps can be seen as the semantic foundation of a conceptual hierarchy. Topological information can be extracted from geometrical maps with various techniques such as so-called fingerprints. A qualitative representation often considered adds *semantic place information* to spatial maps. Places having a direct relation with a human understanding of the world can be learned with supervised learning methods. To provide a common conceptual ground that robots and humans can use to communicate, DFKI [133, 279] introduced a *conceptual layer* to multi-layered maps. Using conceptual abstraction, linguistic reference to spatial areas can be established in HRI [133, 277] (DFKI).

Progress. NIFTi investigates how spatially grounded multi-layered models can be extended to spatio-temporal hierarchies for use in USAR missions (§1.3.5.1,WP1). To arrive at these hierarchies, NIFTi addresses data association problems at a novel level of complexity. NIFTi investigates robust methods to fusing additional mapping modalities, such as vision information, with spatio-temporal representations into hybrid maps (§1.3.5.1,WP1). Methods for concept learning and image understanding (§1.3.5.2,WP2) yield information about objects and events that are salient in disaster environments. NIFTi grounds this information in the spatio-temporal representation of the cognitive architecture, to continually assess a user's cognitive make-up, and improve human-robot interaction and decision making in collaborative planning (§1.3.5.3,WP3; §1.3.5.5, WP5). A central question NIFTi addresses is how to control granularity of the information a robot should convey about spatial structure and dynamics, when communicating.

1.2.3 Observation models for situation awareness

A rescue robot has to be aware of the changing situation around it. Information about the real world comes from onboard sensors, e.g. cameras, laser range scanners, microphones, thermometers, smoke detectors, chemical sensors. The static and dynamic aspects of an environment have to be perceived, particularly landmarks, terrain features needed for robot navigation through passable corridors, unexpected threats, and potential human victims.

We focus here on visual senses. Percepts from sensors have to be preprocessed, and only information relevant to the task is to be provided. Such percepts constitute the lower level in a knowledge representation hierarchy. The upper level acts on a symbolic level. The gap between these levels has to be bridged, by relating relevant percepts and symbols – the symbol grounding problem [83]. Categorical understanding of objects and events reveals more about their functionality, and potential effects especially for USAR environments. The challenge in NIFTi is boosted by the requirement to cope with dynamic and static threats, terrain features, and victims. These events are unexpected from the robot's point of view.

Many projects in mobile robotics have investigated modeling environments. Only a small subset of relevant approaches is mentioned here. The established practice in image/video understanding is to work with intermediate representation levels of objects. The notion of an object allows to focus on percepts relevant to the task performed, and abstract away the background. Object identities can be established by classical image analysis methods based on image/video frame segmentation, description and classification [230]. (We omit a review of the vast literature on this subject for space reasons.) The observations have to be filtered out, clustered (categorized) and qualitative symbolic information derived from them [92].

An important functionality of a rescue robot is to detect common objects which could serve as navigational landmarks, e.g. doors, door handles, windows, distinct pieces of furniture, information labels (e.g., pictographs for exit, toilet, elevator, escape way). State-of-the-art methods can detect distinguished points/regions and find their invariant descriptions [159]. Such methods allow showing objects of interest during the learning phase to the system. In the CTU system, the Maximally Stable Extremal Regions [154] are detected in the image, affine invariant descriptors calculated on them and such entries stored in the database of known objects. In a run mode, the objects are quickly detected in images/video [160].

The other commonly used possibility is to use shape of detected objects. Several approaches have been used to determine the distance between two shapes have been proposed, such as analysis of moments, shock analysis, skeleton analysis. Implicit and Explicit Analysis proposed by ROMA [200] uses the probability density functions as descriptors of datasets and estimates the distance between the two probability distributions obtained as a distance between the two related sets. Their approach can be used to detect human victims from silhouettes. In this field we can distinguish between part-based [161, 146, 276] and global shape based methods [37]. The former have demonstrated more robustness under body occlusions and pose variations; global template matching can be used as a prior to reduce the search space of possible parts combination [146].

Object description is often followed by classification. The state-of-the-art classifiers are comprised in the public domain Statistical Pattern Recognition Toolbox in MATLAB, `http://cmp.felk.cvut.cz/cmp/software/stprtool/`, developed at CTU and used world-wide.

An important functionality for the rescue robot is to track detected objects. Different principles have been suggested. We will concentrate on regression based methods, which do not require any explicit criterion function and estimate the object state directly from the observed intensities by a learned regression mapping. Two important issues for regression methods are modeling of significant changes in appearance, and learning efficiency. Jurie and Dhome [111] suggest to learn a linear mapping between observed image intensities and corresponding motion. The approach can be generalized to non-linear prediction as demonstrated by Williams et al. [270], who learn predictors using a Relevance Vector Machine. CTU colleagues [281] proposed to represent an object by a collection of learned linear predictors. They address the problem of predictor complexity and optimal selection of view-specific set of predictors.

Dynamic threats are unexpected events. Their detection relies on detecting changes in the scene. Motion detection methods [232] provide one such tool. In NIFTi they will be modified for use with a moving camera, processing streams from a moving UGV and/or UAV. Higher level changes will be detected through mismatches between the perceived scene and the higher level representation in a map. Detection of human victims is difficult. The most quick and reliable people detection methods rely on detecting human faces in the upright position. However, the victims often lay or remain in very strange positions. As far as we know, new victim detection methods have to be developed in NIFTi.

The omnidirectional camera can be used for visual orientation of the robot [194]. Recent approaches show that the principle is useful in autonomous driving [48, 85]. The vision subsystem will also help the rescue robot to find flat routes which are easy to navigate. If the correspondences between image entities seen from two or more locations then the homography between image sensor and the world (to be) plane can be calculated. If the homography condition is violated then uneven surface is detected. This is a standard procedure in 3D vision.

Progress. Scientific progress is expected in several areas (WP2, §1.3.5.2). The symbol grounding problem is a challenging one. Most work of others discuss it on the conceptual level. There are only a few papers showing the approach in a very constrained real world. NIFTi sees an opportunity in devising a more layered, hybrid approach, by combining ontology-based information fusion methods with structural pattern recognition methods for objects with strong structure. NIFTi experiments with methods based on 2D grammars [205]. For low-level perceptual grounding, NIFTi investigates object detector/tracker based on the learnable templates. Also, CTU has ongoing and yet unpublished efforts in learnable tracker which represent both object (foreground) and the background. Videos from the stereo vision head rely on good correspondences. We expect to transfer the correspondence seek prone to false alarms for static images [218] to videos from stereo vision head which will be key enabling technology for finding passable terrain. The omnidirectional camera gives a global picture. It is expected to provide for triggering attention mechanism by detecting unexpected events as dynamic threats.

1.2.4 Cognitive robot architectures

Several “cognitive” approaches have recently been proposed for modeling information processing architectures for robots. NIFTi aims to develop cognitive systems that employ insights in how natural cognitive systems can robustly process multi-sensory information at different levels of perception and cognition, and at the same time take advantage of a robot being a computational platform with its own perceptual and processing characteristics.

NIFTi focuses on models for multi-sensory processing, and methods for interconnecting representations across modalities. The latter includes the classical “symbol grounding” problem, connecting perceptual- and conceptual representations to ground content from dialogue and planning, but also connecting cognitive user models with environment models and plans. An important issue is how representations can facilitate information exchange across modalities. NIFTi adopts CAST [86], which permits the specialized representations necessary for integrating the typical processing components used in a robotic system. In CAST, processes interact through working memories and use ontologies and probabilistic models as a basis for exchanging and interconnecting information. This is similar to the mediated models of Steels et al and Roy et al.

DFKI has started developing a probabilistic formulation of [100], using a Bayesian approach. [100] based grounding on the idea that subarchitectures write a-modal representations of content (called proxies) to a working memory, after which these proxies get bound (grounded) into unions. Binding was done on a symbolic feature-by-feature comparison. The Bayesian approach reformulates proxies as content structures with associated hypotheses for how features of that content could be interpreted. This adequately represents the uncertainty in interpretation and allows for the discrete or continuous features to be compared across any number of proxies. Given proxies and their hypotheses, the probability of a union given the evidence of the modal observations of the proxies is the computed with a Bayesian network of feature probabilities. The network specifies the dependencies or correlations between feature instances. This offers a strong theoretical foundation for a robust and adaptive approach to symbol grounding, with several efficient machine learning algorithms in existence for learning both the parameters and the structure of such models. Their graphical nature provides a stronger model for grounding than aforementioned mediation models, as it can in principle capture dependencies over entire structures. To relate structures across different modalities, co-indexation is used (cf. also resolution, §1.3.5.3).

Progress. NIFTi uses CAST as the architectural framework, to integrate spatial- and observation models for modeling the environment at various levels of abstraction (§1.2.1; WPs 1, 2), coupled to multi-modal HRI (§1.2.6, WP3), cognitive user models (§1.2.5, WP4) and flexible planning (§1.2.7, WP 5). The architectural framework thus provides the mindset and technical framework for achieving the NIFTi objectives. The Bayesian network approach to symbol grounding provides a basis for grounding complex conceptual structures. NIFTi further develops and extends this approach, to yield a theory of grounding conceptual information from cognitive user models, planning and communication in functional environment models. Particularly, NIFTi explores extensions to facilitate bi-directional links in grounding, so that user models, and planning- and dialogue-contexts can interact to prime observation, and the use of controllers to online acquire and apply sets of parametrizations. Adaptive controllers guide how and when content can be interconnected across processes, and percolates content changes coherently up and down the architecture. This percolation is based on co-indexation (*that* particular bits of content are related across modalities) and observations on how content can or cannot still be unified on the graph model (detecting changes, and tracking them with a situated history; cf. WP5 §1.3.5.5).

To provide a temporal perspective on how action and interaction are grounded in situation awareness, NIFTi develops a situated exploration history. This history is coupled to DFKI's description logic-based forms of ontological and temporal reasoning [128, 127]. These representations are indexed, ontologically sorted representations akin to modal logic, and are phrased in a decidable fragment. These graphical representations can be directly translated into the type of proxy structures used by the probabilistic approach to binding. Spatial indices are connected to the spatial organization through the conceptual level of representations in the functional environment models. The temporal representation is based on the notion of a time slice. This provides the possibility for dealing with underspecification in the representation of time, as well as an arbitrarily fine granularity of time. This allows for integration with the planning-specific temporal representations discussed in §1.2.7 and §1.3.5.5. Inference over temporal representations is based on Pellet, OWLIM, and Jena, and backed up by Sesame.

1.2.5 Cognitive taskload and selectional attention

In human-computer interaction (HCI) there are different examples of using adaptive systems to enhance operator performance by cognitive taskload reduction (e.g. [187, 170, 267]). Research shows that HCI can be further improved by integrating man and computer symbiotically in a closed-loop system (augmented cognition). The task of the computer is to detect both the user's cognitive state and the operational context to adapt in real time, thereby improving the total human computer performance [134]. In [78] a user model is proposed containing 5 different categories: permanent characteristics (e.g. personality, gender), dynamic characteristics (e.g. fatigue, emotional state), baseline state, momentary state (e.g. which task is in focus at the moment), and the critical taskload model developed by [175, 177, 77]. Most aspects of this user model can be directly transferred to the domain of adaptive HRI. They allow for a more personalised interaction. For example, they can be used as a basis for mitigation strategies to reduce the user's cognitive load [196]. The main difference between the domain wherein this model was applied and the current model is that the user in HRI is not bound to one location. This implies that determination of a task that receives most attention of the user or task switching is much harder, because eye-tracking is too invasive to use in the field. To estimate which tasks are active, possible solutions are to combine the operation plan (including the briefing results), the location and context information, the running commentary and communication patterns, and the direction the human worker moves. For mobile police officers, this makes it feasible to establish adaptive notifications (taking into account officers' cognitive taskload and task priorities).

An additional option is the use of biologically inspired models for attention determination. Biologically inspired models can provide functionality for learning models of human attentive behaviours in circumstances that are similar to the ones NIFTi investigates. The purpose of this kind of attentive approach is to build 3D saliency maps. Research on biological attention has been carried out since (W. James 1890). Location-based and search-based visual attention has been discussed by Treisman and Gelade [252] and Posner [203]. The first ones described the ability to individuate single features as a parallel process (pop-out) while Posner studied the cueing effect, in the presentation of expected and unexpected stimuli. Recently other aspects of attention have been considered, such as change-blindness [227], attentional shifts in 3D scenes [167], top-down components, induced by memory cues or the semantic interpretation of the scene [274]. These studies have inspired computational models for attention and computational architectures amongst which the most well-known are those of [98] and [27]. The first model builds on a parallel computation of individual features further combined in a single saliency map, which is analysed serially. Tsotsos' model, on the other hand, proposes a pyramidal architecture of feed-forward Winner Take All networks, to extract bottom-up features. In both models it is possible to integrate top-down elements to model a saliency outcome on specific tasks. Other computational models include that of Frintrop [66, 65], of Ouerhani [174] for robot navigation. Top-down and bottom-up components and their correlation with depth, ocular movements velocity, fixation clustering and head movements, in scanpath generation have been studied in [15, 14] to model visual scanpath that can be used by autonomous agents to explore a scene. Biologically inspired models for allocation can also reduce the computational workload for the robots.

Progress. NIFTi advances regard two main aspects of attentional models: the first concerns learning salient factors in critical environments that are used by the robot for enhancing tasks abilities (for both execution and switching), and the latter concerns the human factor, via cognitive user models for HRI and decision making processes. The enhancement of task abilities by using salient factors of the environment contribute to both objective 1, contributing to domain specific actions, as 4, resource allocation based on salient objects. The cognitive user models attribute to objective 2 and 3 by enhancing the HRI based on the user state.

The first aspect relies on the works of [10, 12, 1, 9, 13, 14, 15] and on the use of the gaze machine [199, 216], together with motion classification, for building a dynamic model of 3D saliency. The main difference with other biologically inspired models is the combination of dynamic 3D saliency maps with motion classification, which is made possible by the great flexibility of the gaze machine, with respect to usual eye trackers or other eye-movement detectors (see a discussion in [216]). The output of the machine is a 3D dynamic saliency map of presumably attended fixations. Indeed, the map can be realized under several constraints, like distractors, parallel, competing and switching task execution, etc. Constraints are used to capture attended fixations, as for example for localisation and motion classification during quite hard tasks, such as searching for victims signs while traversing troublesome terrains, as it is usual in rescue environments. Understanding and learning localisation and motion features is completely new as it provides fixation maps in which suitable deprivation of sight, during task performance, can effectively identify meaningful fixations during different/parallel and switching tasks performance. These models can be learned by the robot simply by imitation.

On the other hand NIFTi contributes to SOA progress also by decreasing the cognitive task load of the user and the workload of the robot. Here, the aim of the *human factor* is to make the interaction between robot and rescue worker more like between human. The robot is able to detect critical states of the user by combining different characteristics of the user with the context. In the preparation of operations, a work agreement can be set on the adaptation mechanisms (e.g., changes of user-interface mode during the operation), according to the favourable results described in [182]. When a critical state arises, the robot can postpone or change its interaction (voice/content) or its communication style [222, 157, 8], take over tasks, guide the user in performing its task, or request other partners to help (WPs 3–5). This goes beyond the state-of-the-art in that the user model is only validated in indoor environments and measurements used there to disambiguate the information cannot be used in the field.

Finally NIFTi contributes to decreasing the cognitive task load of the user and the workload of the robot by providing user models and biologically inspired attentional models (WP4, §1.3.5.4).

1.2.6 Human-robot interaction

When humans process situated dialogue, they combine information from various sources, e.g. linguistic competence, conceptual “world” knowledge, and spatio-temporal information [118]. Work on human-robot interaction (HRI) is starting to follow up on these insights [214]. Different approaches investigate how situation- and social awareness can be used to extend the notion of *context-sensitive dialogue processing*: dialogue is interpreted relative both to what has been said before (dialogue context), and to situated context. Investigating how to relate dialogue to situation awareness has led to several more general considerations about the nature of situated dialogue processing for HRI. One observation is that dialogue needs to be processed incrementally, if the situated context is to prime comprehension. An *incremental processing model* builds up semantic representations for an utterance in a word-by-word ”left-to-right” fashion, enabling the integration of extra-linguistic information as each word gets analyzed. This can lead to significant improvements in speech recognition and subsequent linguistic interpretation [147, 148].

HRI has been identified as necessary to advance USAR robotics [172]. Interaction with rescue robots currently focuses on GUI-based tele-operation. Empirically, [173] have established that HRI for USAR focuses primarily on joint exploration to improve situation awareness. This poses challenges to grounding situated dialogue. Both robot and human typically have to deal with uncertain and incomplete situation awareness, notably requiring handling of clarification, suggestions and explanations. Clarification occurs in dialogue whenever there is a problem in the interaction, which needs to be resolved before the interaction can continue. So far, strategies for clarification have primarily concerned overcoming speech problems. In HRI strategies for clarifying situation awareness have only been studied to a limited degree (and only in indoor settings).

Progress. NIFTi proposes an approach for HRI to build up a shared understanding of a dynamic environment, through communication about joint exploration (WP3, §1.3.5.3). The approach includes novel mechanisms for (a) using context to robustly process spoken dialogue, (b) user-adaptive production of content to keep up a running commentary, and (c) adaptive resolution, clarification, and grounding of communication. These efforts contribute to achieving Objective 3.

Using CAST (§1.2.4), NIFTi formulates the functionality for multi-modal HRI as a subarchitecture. Within this subarchitecture individual processes capture different aspects of required functionality, e.g. speech recognition or incremental parsing. Processes interact via the subarchitecture's working memory. A working memory is a data storage, containing a collection of partial or complete representations. The representations stored on working memory typically constitute the "current context" in communication, based on foregrounded context and what is currently perceived in the environment. (The latter type of representation is provided by processes external to the HRI subarchitecture.) Processes provide, monitor, and retrieve these representations. A process can thus provide a representation to working memory, which is then retrieved by another process for further processing. Representations on working memory can be linked indexically. This linkage yields a dynamic notion of representation. By linking across representations on working memory, and –possibly– maintaining links to representations on working memories external to the HRI subarchitecture, changes in interpretation can be dynamically percolated throughout an architecture. For communication this is relevant in two ways. First, it provides the basis for context-sensitive processing. Second, as working memory focuses on what is current in the communicative context, it is limited in the spatiotemporal information it contains for reference resolution. Bi-directional connections to longer-term memories are used to resolve large-scale spatial and temporal references, and foreground this content in the HRI working memory.

NIFTi moves beyond the state-of-the-art in speech processing and parsing of spoken dialogue by considering the integration of the these processes together with dialogue interpretation [148], using the possibilities offered by a CAST working memory. Seeing processing of spoken dialogue as an integrated process makes it possible to better deal with typical problems in spoken dialogue such as recognition quality, ungrammatical or incomplete utterances, and pervasive ambiguity. NIFTi combines speech recognition with other levels of processing to help alleviate recognition problems, and improve overall performance. [147] show how speech recognition can be improved by making it context-sensitive. Using information about salient objects, events, and places in the current context, lexical items are activated. These lexical activations in turn balance the language model to "listen to what is most likely to be heard" (while still allowing for backing off to less likely interpretations). [148] propose to use the resulting word lattice directly in a context-sensitive, incremental parsing process. This parsing process combines a linguistic grammar with (a) non-standard rules for combining expressions, and (b) discriminative models for parse selection [280]. The non-standard rules extend coverage to deal with missing words, dysfluencies, discourse-level composition, and correcting typical speech recognition errors. Parse selection models help to select the probabilistically most likely analyses, given the linguistic structure, content and context of the utterance. The parse selection models are learnt using a perceptron model trained on a domain corpus, and are applied after each incremental parsing step. In NIFTi, these initial developments will be developed further. To deal with stress-induced changes in voicing NIFTi intends to use stress detection for adaptive speech recognition [82, 197, 104]. NIFTi proposes a local controller to dynamically control information flow and processing at these different levels of spoken dialogue comprehension. Control policies are learnt off-line from domain corpora (gathered in WP3), and can be adapted online, using a mixture of reinforcement learning and statistical relational learning.

NIFTi advances the integration of cognitive user models into interactive cognitive architectures in several ways. Multi-modal communication comprises advanced spoken dialogue, ecological interfaces [186, 73], and pen-based gesture to aspects of the GUI visualization. The production of spoken dialogue and visual presentation is user-adaptive, using a close coupling between decision procedures in production, models of linguistic cognitive load, and cognitive user models. For comprehension, NIFTi uses mentioned controllers to establish bi-directional links between cognitive user models and stress detection for adaptive speech recognition.

Finally, NIFTi provides a novel approach to "symbol grounding" (cf. §1.2.4; §1.3.5.3, WP3). NIFTi uses a Bayesian network-based approach to move beyond grounding of single categories. It advances on categorical complexity, grounding entire conceptual structures and not just individual concepts, in combination with a layered approach to perceptual grounding (cf. §1.2.3; WP2). Furthermore, it grounds concepts in space and time, and can track changes over space and time in conjunction with the situated exploration history (§1.3.5.5, WP5). This provides HRI the possibility of communication to reference long-term temporal perspectives. Finally, NIFTi adds mechanisms to integrate reference resolution and referential aliasing directly into the grounding process. Grounding includes clarification as an explicit mechanism in grounding, to resolve uncertainty and incompleteness.

1.2.7 Cognitive execution, learning and flexible behaviors planning

Robot cognitive execution is highlighted in many biologically inspired architectures, like ISAC [114], the ALEC architecture based on state changes induced by homeostatic variables [67], Hammer [43] and the GWT (Global Workspace Theory) based cognitive architecture [224]. In attention-based approaches, e.g. [113, 55, 174, 65, 15, 43, 32] cognitive execution is induced by selective attention to filter out distractors and maintain stimulus priorities [61]. Attention based-control has been considered mainly for feature extraction in self-localization and topological navigation such, e.g [71, 272, 174, 93, 2]. While attention mechanisms act on the perceptual stimuli, it seems that even if stimuli are the same for two tasks there exist executive control processes intentionally preparing to switch cognitive "sets" to perform one or the other task [62].

The neuroscience studies on cognitive execution, and mainly on human adaptive behaviors suggest to model robotics behaviors via inhibition-based executive processes. Theories on executive control processes and task switching have strongly influenced cognitive robotics architectures since the eighties, e.g. the Norman and Shallice [188] ATA schema, Duncan FLE model[47] and the principles of goal directed behaviors in Newell [183].

In the approaches to model based executive robot control, tasks and activities are general robot modes and the runtime system manages *backward inhibition* via real-time selection, execution and actions guiding, by hacking behaviors [19] considering different modalities [107]. From these ideas the Constraint Based Interval Planning (CBI) paradigm was introduced, showing a strong practical impact when it comes to real world applications, see e.g. RAX [110], IxTeT [69], INOVA [244], and RMPL [269], and also the ASE system, further integrated in CLARAty (see e.g. [54]). Dynamic allocation of resources in multi-robot coordination has been investigated in [143], proposing a taxonomy on robot capabilities to perform a single or many tasks at a time, and on task allocations. [223] have addressed the need to obtain dynamic prediction of task duration given the current robot team state, and similarly in [120, 121] constraint optimization has been used to handle coordination during communication breakdowns. Only recently cognitive control based on learning from experience has been investigated with Robel [166]. On the other hand, in the theory of action and change framework the problem of executive control has been regarded mainly, both for off-line and online action execution, in terms action properties. Actions effects on the world and the agent's ability to decide on a successful sequence basing on its desire, intentions and knowledge [144, 3, 145], have been represented from the stand point of their ontologies. In this sense cognitive execution is intended as the reasoning process underlying the choice of actions.

Progress. NIFTi incorporates new ideas to address the dual interaction problem arising from human-robot cooperation, multi-agent learning including models of other agent behaviors, reasoning support on choices performed and task-driven attention (§1.3.5.4, WP4; §1.3.5.5, WP5). These are given in a logical framework ensuring coherence of multi-task preconditions and effects, allowing the agent team to reason about decisions and choices made, while maintaining flexibility of plans [57, 32, 31] (WP5). Though necessary, this is not sufficient, because constraint based techniques, lacking learning principles, fail to incorporate a notion of the effects of team-mate actions (both human and robot) in the development of strategies and policies and do not consider flexible and adaptable behaviors that can arise from human-robot cooperation or robot-environment interaction. Therefore a further crucial point to investigate is multi-agent learning, incorporating models of other agents behaviors, including the environment reaction to the agent actions. Studies on multi-agent learning have moved the area of learning in dynamic environment from reinforcement learning to game theory [75], and online estimation of Markovian system dynamics [207] to determine, from observations, the behavioral trends of the environment through interaction. Though there are several trends on multi-agent learning from observations, those inspired by the developmental learning principles are the most appropriate for cognitive execution based on context adaptation and task-driven attention [9, 15, 1]. Imitation of actions and action-reaction based behaviors seem to be the most appropriate to face the complex hierarchy of skill learning in order to coherently interface all the decision steps to link task choice, task switching and action execution. From the earliest studies on action imitation [248] and [33] NIFTi makes progress on new methodologies for learning simple robot behaviors [16, 14] and on action-reaction learning [106]. NIFTi extends these principles of developmental learning to properly construct a mapping from internal states to execution control for tasks choices under time constraints and compatibilities. A NIFTi cognitive architecture acquires new skills through exploring new environments, and jointly acting and interacting with humans. A cognitive architecture properly learns (and where morphologically possible, imitates) afforded actions (WP5) and ways of communicating (WP3, §1.3.5.3), by observing salient aspects of action and interaction. Together, this contributes to achieving Objectives 4, and 3.

1.2.8 Adaptive robot morphology

The scene after a disaster is dangerous for a rescue team and needs to be rapidly secured before operation. This justifies the use of robots as it makes the search for survivors safer for rescuers, and robots can be deployed immediately. [35] present a detailed description of using robots in the NYC World Trade Center. [172] summarizes the conclusions from that experience. [172]'s main conclusion is: *the existing approaches to locomotion are too limited.* Although robots could inspect places (e.g. narrow tunnels) unreachable for dogs, humans or tethered cameras, they often failed to cross obstacles. In USAR, robots need to demonstrate strong off-road capabilities. Most locomotion concepts are based on wheels, caterpillars, legged or snake-like robots. We discuss these concepts in turn, and analyze trade-offs in developing suitable robots.

Walking machines (e.g. Dante [7]) are well-adapted to unstructured environments as they can ensure stability in a wide range of situations, but are mechanically complex and require a lot of control resources. On a planar surface, they have a low speed but high power consumption in comparison to other concepts. *Caterpillar* robots demonstrate good off-road abilities thanks to their stability and good friction (large footprint), but suffer high friction losses between surface and caterpillars while turning. Caterpillars offer good mechanical robustness and a large footprint, allowing the robot to move on various types of soils (e.g. sand, mud, rocks). The other advantage of the large footprint is that it allows to pass over relatively larger holes. Most commercial rescue robots have caterpillars and are relatively heavy. Examples of such robots are TeleMAX[1], PackBot[2] and LMA-1[3]. *Wheeled rovers* are optimal for well-structured environments, e.g. roads or buildings. Off-road, their efficiency depends on the typical size of encountered obstacles to be crossed. This holds for Sojourner [236], Rocky 7 [265], the Volksbot family [241], or Micro5 [135]. They can typically cross obstacles due to their wheel size, if friction is high enough. Using parallel structures, it is possible to improve the climbing performance of such rovers. [226] presents Shrimp, which can climb steps up to two times the rover's wheel diameter. Further increasing climbing capabilities of a wheeled rover requires the use of a special strategy and often implies dedicated actuators like for the Marsokhod [117] and Hybtor [142] or complex control procedure like for the SpaceCat [138], Nanorover [258] or the Octopus [139]. Finally, *snake-like robots* present the best

[1] http://www.telerob.de
[2] http://www.irobot.com
[3] http://www.esit.com

climbing capabilities, being well-adapted to traverse extremely difficult terrain. An example for USAR is the OmniTread [4], which combines a serpentine locomotion mode with caterpillars. The main drawbacks are the complicated control, high power consumption and slow motion.

Locomotion concepts are usually classified into active and passive locomotion. Passive locomotion uses passive suspensions, having no sensors or additional actuators to guarantee stable movement (e.g. Shrimp), whereas an active robot employs a close control loop to maintain stability during motion (e.g. Octopus). Active locomotion extends the mobility of a robot but increases system complexity and needs extended control resources. Increasing the number of active degrees of freedom has a negative impact on different crucial aspects i.e. sensing, control complexity, weight, energy and reliability [53, 273]. This suggests a trade-off between the two structure types. A hybrid approach integrates both active and passive adaptation mechanisms. and is the best performing structure if we consider all the criteria together (i.e. power, complexity, climbing etc.).

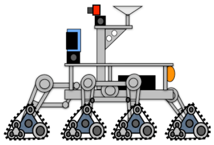
Progress. By analyzing the market, mainly two families of rovers can be discerned: military rovers, and space rovers. Basically, there are no rovers with a dedicated USAR focus and use, as only military rovers are commercially available. While military rovers are mainly completely remote controlled, active and highly robust, they do not focus on autonomy. On the other hand, space rovers are almost exclusively passive (structure), they implement some autonomy, and are highly optimized in terms of weight and energy. NIFTi has the aim to improve on the current systems by analyzing the needs (with the end-users), and by combining active technologies permitting variable morphology and passive space like solutions fro high climbing ability.

Most commercially available platforms currently used for USAR applications use a purely active design with two pairs of caterpillars. The angle between the pairs of caterpillars can be modified through additional motorized degrees-of-freedom. Such an active variable morphology allows to improve the climbing capability of the robot. Even though such robots can be used in the context of USAR, they still lack sufficient climbing capabilities and autonomy.

NIFTi proposes to use space robotics technologies robotics and BLUE's experience in this field to design a locomotion concept that outperforms existing ones (WP6, §1.3.5.6). The constraints on weight, reliability and power are very strong for space applications, yielding rovers which are light, simple and have a low power consumption for their operation. All these constraints favor locomotion concepts with wheels and passive adaptation mechanisms. Such rovers offer a good climbing capability while keeping control simple thanks to the passive adaptation. The integration of space robotics concepts into the USAR domain presents a great potential for improving USAR robots. It will allow us to optimize for weight, power consumption, control complexity and climbing performance. Figure 4 summarizes our approach. The development of the platform is user-centered. End user requirements are integrated at an early stage of the design phase, to guarantee NIFTi obtains the best suited platform and avoid too complex or unusable solutions.

Beyond morphology adaptation for locomotion, NIFTi also investigates adaptation for perception. NIFTi develops methods for using a UAV as roving sensor for the robot (WP6, §1.3.5.6) [81, 239]. This form of active perception moves beyond the active use of different perceptual modalities (stereo- and omni-vision, range finders) which are in a fixed configuration on the robot platform. The robot's cognitive architecture can decide to deploy the UAV to provide visual data for aspects of the terrain which are currently invisible or occluded to the robot's fixed perceptual modalities. The cognitive architecture makes these decisions on the basis of currently available terrain information, anticipations about what lies beyond the currently visible, and trajectories for its exploration plan (WP5, §1.3.5.5; WP6, §1.3.5.6). This contributes to Objective 4.

[4] http://www.engin.umich.edu/research/mrl/00MoRob_6.html



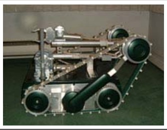
ESI – LMA 1	Innovative concept	BlueBotics - Shrimp
	Innovation takes the best of the two worlds	
Fully active design	Hybrid architecture	Passive design
x complex control x active adaptation to terrain (slow) x heavy construction x high power consumption ✓ large footprint ✓ active dof helps to climb tricky obstacles	✓ optimized weight ✓ adaptive concept ✓ high climbing performance ✓ simple ✓ reliable	x wheels can get stuck in concavities ✓ low power consumption ✓ simple control ✓ light-weight ✓ intrinsic self-adaptation to terrain

Figure 4: The new locomotion concept (left) merges advantages of active, passive designs (right)

1.3 S&T methodology and associated workplan

1.3.1 Overall strategy and general description

The work plan follows the NIFTi roadmap (§1.1.5). The roadmap defines milestones for achieving the project objectives. These milestones formulate increasingly more sophisticated *integrated theories and systems* with specific, measurable capabilities – underlining NIFTi's strong focus on integration. The development towards these milestones is iterative.

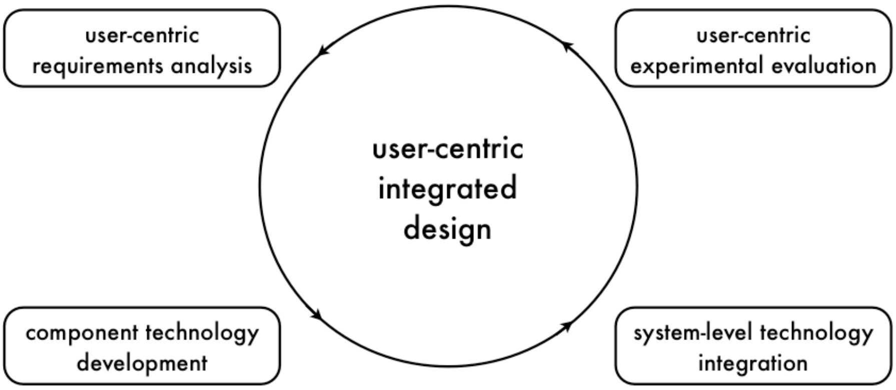


Figure 5: The NIFTi development cycle

Figure 5 shows the NIFTi development cycle: an iterative, incremental process in which design specifications and implementation are successively extended, refined and validated. The design rationale with its claims on human-robot performance will be systematically tested and refined for all components in so-called formative evaluations, such as cognitive walkthroughs with end-users and their representatives, and reviews by Human Factors experts. Yearly, a summative evaluation with the end-users will show how far the general claims are met for the integrated prototype. Furthermore, this evaluation will also provide further insight into the end-user needs, which will be further addressed in the subsequent development processes. Figures 6 (p.33, PERT) and 7 (p.36, GANTT) graphically present the work defined in §§1.3.5.1–1.3.5.7. Figure 6 shows the different work packages, and how their interaction contributes to achieving the project objectives.

WPs 1–6 all focus on developing theories and system functionality *from an integrated, system-level viewpoint*. Situation awareness is *for something*, interaction and action are *situated* and connected to *cognitive user models*, joint exploration needs to *fit in with what humans do and how, and what is considered relevant in an environment*, variable morphology serves a purpose of *efficiently executing actions*. All this functionality is adaptive, to deal with changing uses in novel situations – and *achieves that adaptivity by integrating information from multiple sources*. NIFTi therefore has dedicated a work package to focus on system integration and -evaluation: WP7. The development cycle is fundamental to achieve integration, and to do so iteratively. WPs 1–6

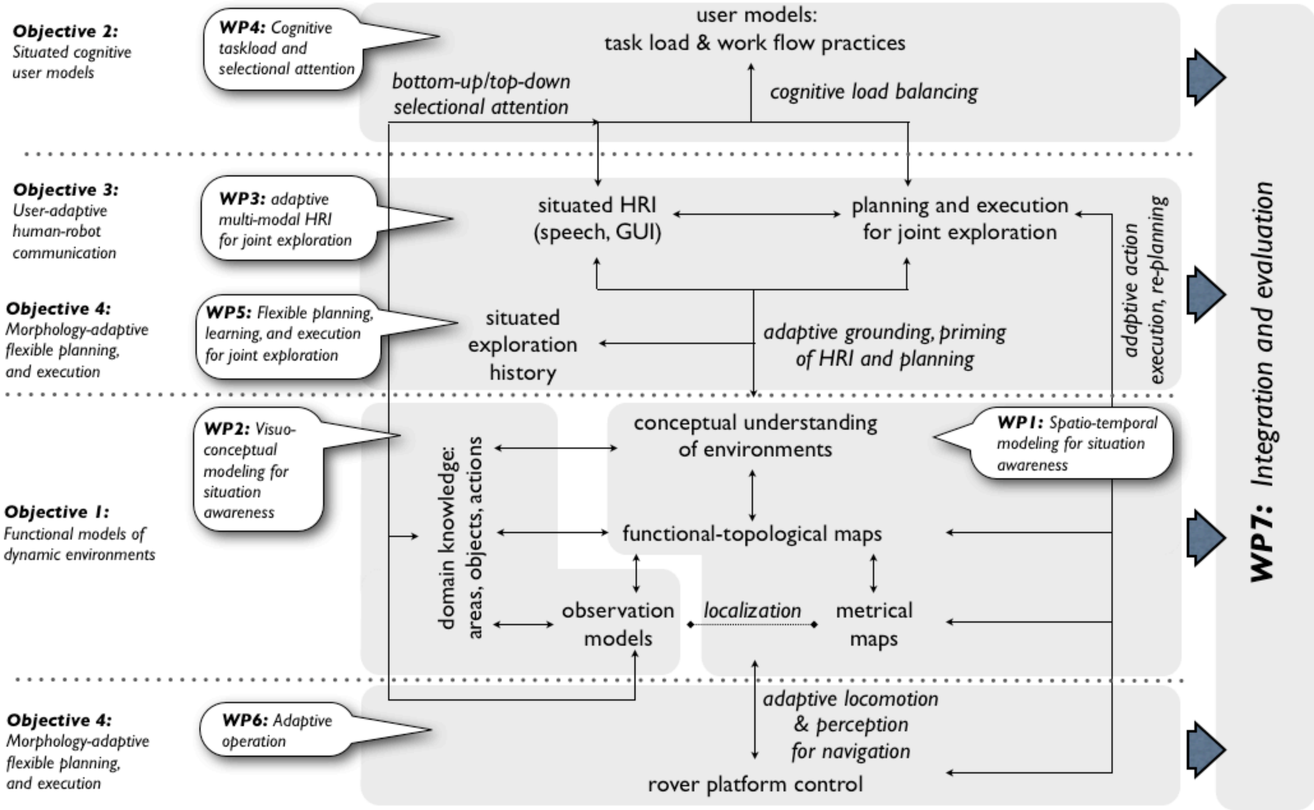


Figure 6: Graphical presentation of WPs contributing to architecture design, objectives

develop individual components. Integration starts at WP-level, as functionality typically requires interaction with components from different WPs. WP prototypes are hence directly provided to WP7, contributing to P deliverables in WP7. NIFTi aims at achieving full integration at system-level every 4 months, in WP7. Component- and system-evaluation occur in-project on an ongoing basis, using agreed-upon unit tests for regression testing. Use case definition and evaluation with end-users are on a yearly basis. The results of domain analyses and evaluations feed back into research, and may be used to adapt the roadmap.

In the work package descriptions, each WP describes a small set of tasks, and milestones. The tasks provide *functionality* which is bundled into *capabilities* specified by yearly work package milestones. The work package milestones in turn contribute to achieving the project-wide milestones defined by the roadmap (§1.1.5, §1.3.7). Figure 7 shows the Gantt chart.

Risk management and contingency planning

Naturally, there are risks associated with the NIFTi objectives. The project provides three means to timely *identify* potential risks: The risk- and contingency plans for the individual work packages, the project development cycle, and management (Executive Board; EXB). The EXB monitors progress in the project, and synchronizes efforts, on a monthly basis through the EXB meetings. This helps us to quickly identify and mitigate any problems that may arise, at the level of the individual work packages. Furthermore, the project's development cycle enables frequent system-level integration, and provides a "continuous" testing of individual modules against unit tests. Regression testing of system-level integration thus gives us further insights in the progress made in modules developed by individual work packages, as well as in their integration in the overall architecture. Finally, NIFTi organizes bi-annual General Assembly meetings, and frequent (at least annual) system evaluations, which further help to establish whether the project is synchronized to its roadmap.

Should any risks be identified, we will aim to *handle* them as follows. Firstly, we will put in place the work package contingency plans to mitigate the risk. Secondly, to ensure that system-level integration is not compromised, we can further allow a rollback to previous results to ensure we keep a stable basis for further development. Thirdly, sub-projects can be established within the project, which can be defined such as to provide dedicated efforts to resolve the identified problems. These sub-projects will be of a limited duration (<3 months) so as not to impede general progress of the project.

To recapitulate, the management procedures, work package risk- and contingency plans, and development cycles, coupled to clearly identified evaluation criteria on the roadmap, all lead to timely checks on progress, which help us to quickly identify any risks that may arise. The above procedures cannot ameliorate every possible risk, but we hope they present a sensible strategy to ensure that overall project progress is not compromised in the long run.

Approved by EC on 2009-12-09

1.3.2 Work package list

WP#	Work package title	Activ.	Lead partic.#	Person months	Start month	End Month
WP1	Spatio-temporal modeling for situation awareness	RTD	5	108.25	1	48
WP2	Visuo-conceptual modeling for situation awareness	RTD	6	100	1	48
WP3	Adaptive multi-modal HRI for joint exploration	RTD	1	92.25	1	48
WP4	Task load and selectional attention	RTD	2	101.50	1	48
WP5	Flexible planning, learning and execution for joint exploration	RTD	7	85	1	48
WP6	Adaptive operation	RTD	4	73.5	1	48
WP7	Integration & Evaluation	RTD	3	127	1	48
WP8	Dissemination and community building	OTH	1	20	1	48
WP9	Management	MGT	1	23.30	1	48
	Total:			730.80		

1.3.3 Timing of work packages and their components

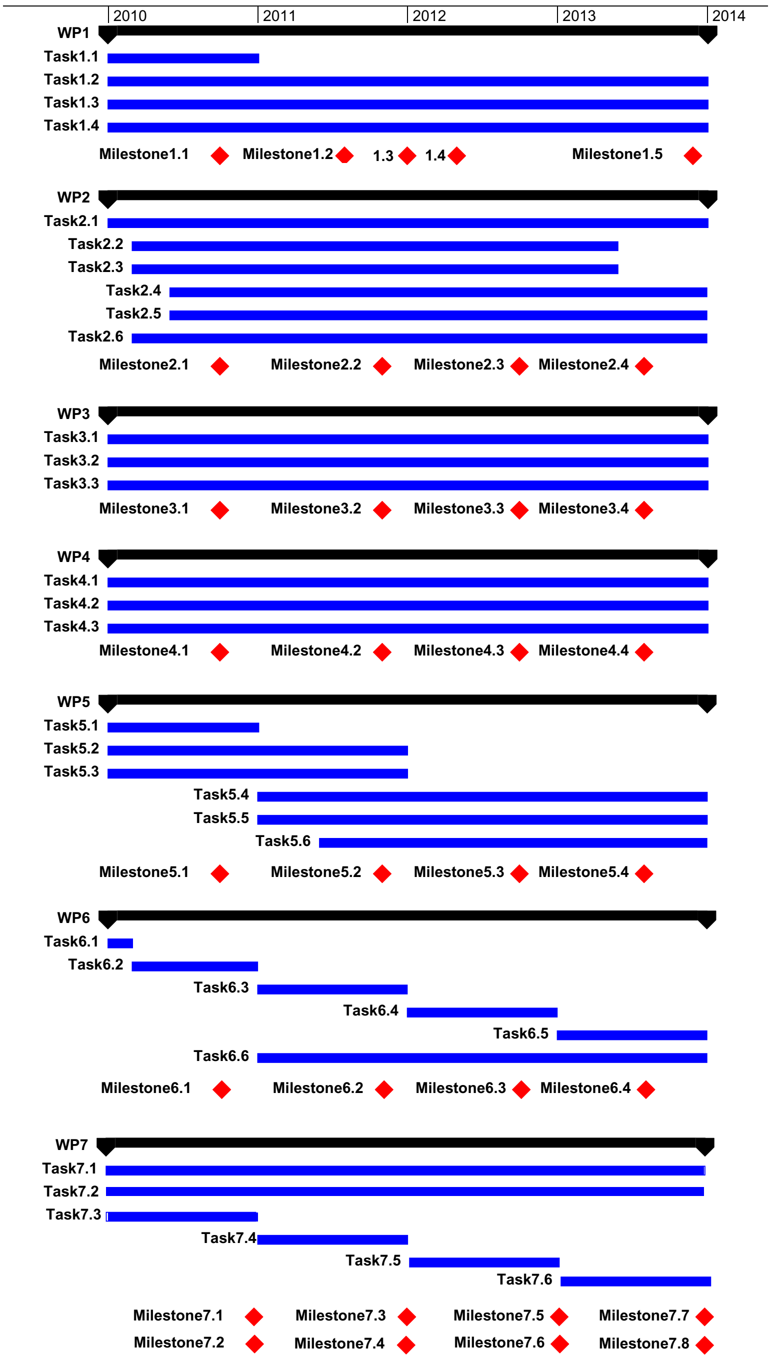


Figure 7: GANTT: Timeline of WP tasks and milestones.

WP tasks are evaluated at yearly WP milestones. Milestones for WPs 1–6 (at month Yr+10) feed, through integration, into achieving milestones for WP7 (at month Yr+12).

1.3.4 Deliverables list

Only WPs 6 (adaptive operation) and 7 (integration & evaluation) have prototype deliverables (P). WPs 1–5 contribute components directly to WP7, to strengthen project-wide integration.

Del.no.	Deliverable name	WP no.	Lead benef.	Est. indic. PMs	Ntr	Diss. level	Deliv. date
DR 7.1.1	Specification of project software development and -documentation standards	7	Fraunhofer	4	R	PU	M3
DR 6.1.1	Platform specification and design	6	BLUE	4.5	R	PU	M6
DR 7.1.2	Specification of current human factor knowledge of the USAR domain and known metrics and collaborative tools in this domain	7	Fraunhofer	3	R	PU	M6
DR 5.1.1	Domain analysis and specifications: context scenario and skills primitives	5	ROMA	9	R	PU	M10
DR 1.1.1	Acquisition of spatial maps of semi-structured environments	1	ETHZ	36.25	R	PU	M12
DR 2.1.1	Basic vision-based situation awareness capabilities	2	CTU	23	R	PU	M12
DR 3.1.1	Adaptive situated HRI for human-instructed navigation	3	DFKI	23.25	R	PU	M12
DR 5.1.2	Methods and paradigms for skill learning based on affordances and action-reaction observation	5	ROMA	11	R	PU	M12
DR 6.1.2	Platform manufacturing and sensor integration	6	BLUE	17	P, R	PU	M12
DR 7.1.3	Integration & end-user evaluation for human-instructed exploration	7	Fraunhofer	22.5	P,R	PU	M12
	Total Yr1:			153.5			
DR 4.2.1	Validated task load, attention and user model patterns that specify the relationship between task demands and user properties	4	TNO	29.5	R	PU	M16
DR 1.2.2	Acquisition of spatio-temporal maps and place topologies of semi-structured environments	1	ETHZ	24	R	PU	M24
DR 2.2.2	Stereo- and omni-directional vision for human assisted exploration	2	CTU	23	R	PU	M24
DR 3.2.2	Adaptive situated HRI for human-assisted navigation	3	DFKI	23	R	PU	M24
DR 5.2.3	Hierarchical structure of learned skills, scan paths, saliency map of activities and communication interfaces	5	ROMA	15	R	PU	M24
DR 6.2.3	Trajectory analysis: principle and evaluation	6	BLUE	13	R	PU	M24
DR 7.2.4	Integration & end-user evaluation for human-assisted exploration	7	Fraunhofer	31.5	P,R	PU	M24
	Total Yr2:			159			

Del.no.	Deliverable name	WP no.	Lead benef.	Est. indic. PMs	Ntr	Diss. level	Deliv. date
DR 5.3.4	Resources management and mapping from states to execution, integrating linear dynamic models learning into theory of actions	5	ROMA	18	R	PU	M30
DR 4.3.2	Theory and evaluation of working agreement method and HRI-adaptation to different contexts based on adaptation guidelines, adaptive IU and interaction model	4	TNO	37	R	PU	M34
DR 1.3.3	Acquisition of human-compatible hybrid maps of semi-structured environments	1	ETHZ	27	R	PU	M36
DR 2.3.3	Bi-directional cooperation of low-level vision modules and higher level control	2	CTU	25	R	PU	M36
DR 3.3.3	Adaptive situated HRI for in-field exploration planning	3	DFKI	23	R	PU	M36
DR 5.3.5	Flexible planning with time constraints and compatibilities	5	ROMA	13	R	PU	M36
DR 6.3.4	User interaction and trajectory planning in unstructured environment based on 3D perceptual data: principle and evaluation	6	BLUE	11	R	PU	M36
DR 7.3.5	Integration & end-user evaluation for in-field joint exploration planning	7	Fraunhofer	32.5	P,R	PU	M36
	Total Yr3:			186.5			
DR 4.4.3	Validated intelligent working agreement mechanism to set-up adaptive HRI	4	TNO	15	R	PU	M40
DR 1.4.4	Spatio-temporally grounded situation awareness using a-priori information	1	ETHZ	21	R	PU	M48
DR 2.4.4	Scaling the functionalities of vision subsystem with the complexity of the environment	2	CTU	29	R	PU	M48
DR 3.4.4	Adaptive situated HRI for in-field joint exploration	3	DFKI	23	R	PU	M48
DR 4.4.4	Summative evaluation and theory of the setting-up and usage of adaptive HRI	4	TNO	20	R	PU	M48
DR 5.4.6	Mixed initiative planning and user requests subsumption. Adaptable strategies for complying to robot-team and users requests.	5	ROMA	19	R	PU	M48
DR 6.4.5	Trajectory planning in dynamic unstructured environment based on 3D perceptual data: principle and evaluation	6	BLUE	28	R	PU	M48
DR 7.4.6	Integration and end-user evaluation for sharing situation awareness	7	Fraunhofer	33.5	P,R	PU	M48
	Total Yr4:			188.5			
	Total Yr1-Yr4:			687.5			

Deliverables for WPs 8 (dissemination) and 9 (management) are listed here for completeness.

Del.no.	**Deliverable name**	**WP no.**	**Lead benef.**	**Est. indic. PMs**	**Ntr**	**Diss. level**	**Deliv. date**
DR 8.1.1	NIFTi project portal	8	DFKI	2	P	PU	M3
DR 8.1.2	Market analysis for USAR robots with HRI	8	DFKI	2	R	PU	M10
DR 8.1.3	Proceedings of the NIFTi summer school Yr1	8	DFKI	3	R	PU	M12
DR 9.1.1	NIFTi annual progress report Yr1	9	DFKI	5.40	R	PU	M12
DR 8.2.4	Proceedings of the NIFTi summer school Yr2	8	TNO	3	R	PU	M24
DR 9.2.2	NIFTi annual progress report Yr2	9	DFKI	5.30	R	PU	M24
DR 8.3.5	Proceedings of the NIFTi summer school Yr3	8	CTU	3	R	PU	M36
DR 9.3.3	NIFTi annual progress report Yr3	9	DFKI	6.30	R	PU	M36
DR 8.4.6	Updated market analysis for USAR robots with HRI	8	DFKI	1	R	PU	M44
DR 8.4.7	Proceedings of the NIFTi summer school Yr4	8	ETHZ	3	R	PU	M48
DR 8.4.8	Public release of the open source NIFTi software	8	DFKI	3	R	PU	M48
DR 9.4.4	NIFTi annual progress report Yr4	9	DFKI	6.30	R	PU	M48
	Total Yr1-Yr4 (MGT+OTH): Total Yr1-Yr4 (MGT+OTH+RTD):			43.30 730.80			

1.3.5 Description of work packages

1.3.5.1 WP1: Spatio-temporal modeling for situation awareness .

Work package number:	1	Starting date or starting event							1
Work package title	Spatio-temporal modeling for situation awareness								
Activity type	RTD								
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	8	0	22	2	62	8	6	0.25	0

Objectives

WP1 delivers functionality for understanding an environment in a spatiotemporal, functional sense. WP1 contributes to achieving Objective 1. WP1 addresses:

(A) Hybrid mapping strategies which fuse appearance-based (visual) mapping modalities with spatial world representations.

(B) Spatio-temporal world representations that cover the environment's spatial structure along with eventual changes of that structure.

(C) Conceptual world representations with a spatio-temporal grounding.

Description of work – Tasks

WP1 focuses on the creation of multi-layered and hybrid maps with a spatio-temporal grounding. At several stages of the mapping process, it is essential to solve challenging data association between map layers and between maps and sensory data. The data association problem is thus the core challenge of WP1, as reflected in the tasks below.

Task T1.1: *Multimodal localization in semi-structured environments [30.25 PM]*
(Period: M0-M12, Partners: [ETHZ/18 PM, Fraunhofer/6 PM, CTU/4 PM, BLUE/2 PM, FDDo/.25PM], Contributes to objectives: A, B)

In general, the mapping problem strongly relates to two essential robotic problems: self-localizing the robotic platforms and consistently associating sensor data with internal representations. In the initial phase of the project, T1.1 refines and fuses well-established localization techniques based on visual and range observations. This enables the acquisition of consistent maps of USAR disaster sites and is the ground for the perception level of situation awareness.

Task T1.2: *Hybrid mapping of semi-structured environments [25 PM]*
(Period: M0-M48, Partners: [ETHZ/18 PM, Fraunhofer/3 PM, CTU/4 PM], Contributes to objectives: A)

Task T1.2 aims at the development and integration of approaches to multimodal robot mapping. From initial domain analyses, we know that we face semi-structured spaces with a sparsity of conceptual information. We expect common structural features, e.g. walls, doors, and furniture to mostly be displaced, damaged, or destroyed. Thus, previously available information may no longer be consistent with the actual state of the environment. The lack of reliable structural information in a disaster scenario demands additional mapping modalities in order to maximize situation awareness.

T1.2 develops robust methods for fusing additional mapping modalities with spatial representations into hybrid maps. T1.2 provides the spatial layers for hybrid maps (see figure below, p.43). T1.2 starts from point clouds acquired by laser range sensors and enriches the maps with visual information originating from various vision sensor setups [123], such as cameras carried by the UAVs, satellite views, and cameras mounted on the UGVs (WP2). In T1.2, the data association problem consists in estimating a comprehensive, i.e. a fully textured spatial model from a network of sensors. T1.2 investigates novel approaches to solving data association from conditionally independent, multi-modal sensor sources, optimally fusing ground based and aerial observations in a principled manner.

Task T1.3: *Spatio-temporal mapping of semi-structured environments [28 PM]*
(Period: M0-M48, Partners: [ETHZ/18 PM, Fraunhofer/8 PM, Rome/2 PM], Contributes to objectives: B)

T1.3 proposes a representation that covers changes in spatial properties of the world along a fixed timeline. T1.3 assumes dynamic characteristics to be a key aspect of disaster environments. In an ongoing disaster, changes in the environment need to be understood by the robot. Also, prior information on the state of the environment can help detect changes that happened beforehand but were not tracked by observation, e.g. the collapse of a building in an earthquake.

T1.3 builds and maintains consistent spatio-temporal representations of disaster sites based on 3D laser sensory readings. It provides the spatio-temporal layer for hybrid maps (see p.43). Extending the probabilistic change detection techniques of [112], T1.3 annotates the environment with change parameters at the point cloud level. It does not explicitly decompose the world representation into static and dynamic entities and it shows that this is a reasonable assumption: Especially in a disaster situation, objects which are usually considered static may eventually change over time, e.g. walls may collapse. Due to this notion of object variability, the complexity of the data association problems encountered within T1.3 is increased with respect to T1.2 as well as to what is generally treated in the state of the art.

Given the assumption of structural sparsity, data association cannot rely on state-of-the-art techniques. Existing methods either apply strong priors about world decomposition or suffer from complexity explosion and aliasing [39, 231]. To maintain generality, T1.3 aims for approaches to annotating spatial maps with change parameters whilst relaxing specific assumptions, such as world semantics or even object identity. T1.3 does put an assumption on how the “scale” of an eventual change relates to the parts of the environment that are left unchanged at a given time e.g. ceiling features are very robust against dynamics and are easy to detect with 3D laser scanner [140]. Such features are good candidates for precise localization and data association.

Furthermore, T1.3 explores possibilities to detect changes between a-priori map information and the actually observed state of the environment. NIFTi believes that appropriate data sources will become increasingly established [219], and will benefit situation awareness.

Task T1.4: *Conceptual functional mapping of semi-structured environments [25 PM]*
(Period: M0-M48, Partners: [ETHZ/8 PM, Fraunhofer/5 PM, DFKI/8 PM, Rome/4 PM], Contributes to objectives: C)

NIFTi uses the spatio-temporal models of T1.2–T1.3 as the physical grounding for situation awareness. They provide the basis of a multi-layered stack of higher-level representations, e.g., topological and conceptual maps. It has been shown that humans cluster space into (functional) regions – a principle that is even applied to outdoor environments without objectively perceivable or physical boundaries [89]. Current approaches to conceptual (indoor) mapping for robots take this into account. In order to become useful for outdoor robots or robots operating in semi-structured environments – and especially to become useful for effective communication with a human operator – a functional understanding of space must be combined with a spatio-temporal environment representation (cf. §1.2.2). To bridge the gap between human-centric and robot-centric representation, the cognitive architecture applies conceptual inference to these representations to infer functionality for changed areas and objects, and projects this functionality into space by representing it as areas in which the functionality is available. This provides the basis for linking cognitive user models and plans to the architecture's environment models.

T1.4 proposes formalisms for grounding topological descriptions and conceptual descriptions in spatio-temporal maps (cf. the figure on p.43). Extending [262], [133, 279] (DFKI), T1.4 identifies new approaches to human-compatible categorization and conceptualization of areas grounded in environment models. T1.4 solves data association through inference: Conceptual descriptors and topological place graphs are instantiated by reasoning over spatio-temporal characteristics of the world. Simultaneously, these descriptors and place graphs are associated with the corresponding spatio-temporal map entities. The novelty is in introducing temporal information into the inference process. Conceptual representations serve to generate linguistic descriptions from spatio-temporal models, and annotate these models through resolution of linguistic descriptions [278]. Reference to map entities, e.g. areas, regions, zones, individual objects, or clusters of objects are established on the basis of semantic labels assigned by both robot and human. The close coupling of spatio-temporal and conceptual information enables an abstract sharing of situational information (WPs 3,5).

Description of work – Milestones

In the progress of the project, WP1 evolves and populates the multi-layered abstraction hierarchy of world representations which ranges from pure spatial data to dynamic, landmark, topological, semantic, and conceptual descriptors.

Milestone MS1.1: *Consistent spatial representations of USAR sites* (M10)
At the end of year 1, the work package goal is to perform human-instructed acquisition of consistent spatial representations of USAR sites. By analyzing variance features of point distributions being implicit within spatial maps, we provide traversability information to adaptive locomotion modules. Since the mapping problem strongly relates to the problem of self-localizing robotic platforms, WP1 implements localization techniques based on visual and range observations.

Milestone MS1.2: *Consistent spatio-temporal representations of USAR sites* (M22)
WP1 identifies an approach to spatio-temporal modeling under the minimization of semantic and object identity assumptions. WP1 consistently acquires spatio-temporal representations covering eventual changes of USAR sites, e.g. appearance, disappearance, or deformation of objects.

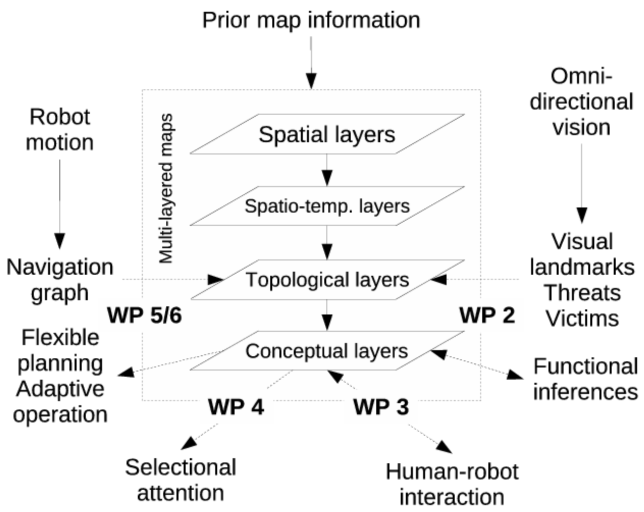
Milestone MS1.3: *Spatio-temporally grounded place topologies of USAR sites* (M22)
Focusing on the topological layers, year 2 is used to enrich qualitative descriptions and annotations of spatio-temporal entities. WP1 infers place topologies over spatial world models, and WP1 explicitly establishes appropriate links between spatio-temporal and topological layers. Perception modules coherently provide and integrate new information, and enable other modules to pull knowledge at different levels of abstraction and granularity.

Milestone MS1.4: *Consistent hybrid representations of USAR sites* (34)
In year 3, WP1 further refines the multi-layered framework in order to comprehend differing perspectives of spatial and spatio-temporal landmarks. WP1 fuses visual mapping modalities into the spatial representations. WP1 deploys UAVs and UGVs in large-scale USAR domains and consistently acquires hybrid maps at a novel stage of complexity.

Milestone MS1.5: *Spatio-temporally grounded conceptual descriptions of USAR sites* (M44)
Starting from the third year, WP1 extensively works on populating the upper, the conceptual hierarchy. WP1 infers conceptual descriptions for spatio-temporal world models. Also, WP1 establishes a grounding of concept descriptors in the spatio-temporal hierarchy in order to facilitate shared SA between humans and robots and collaborative action planning (with WPs 2,3).

Description of work – Contributions to project & state-of-the-art

The diagram below depicts the integration interdependencies of WP1 functionalities with the functionalities of other WPs. Outbound arrows represent project contributions.



The spatio-temporal models resulting from WP1 serve as the physical grounding of higher level concepts from WPs 2–5. Visual landmarks obtained in WP2 are anchored in the WP1 spatio-temporal maps and introduced into the topological and conceptual reasoning process. To WP3, a conceptual understanding of the environment becomes highly relevant when planning dialogues and interactive flexible behavior. WP5 makes use of spatio-temporal and functional information to guide skill learning. Cognitive execution elaborates on a proper mapping from internal spatio-temporal states to tasks, to build time-flexible behaviors. WP6 relies on spatio-temporal and topological representations as the basis for adaptive robot control and robot navigation. Selectional attention mechanisms from WPs 4–5 are driven by spatio-temporal and topological model entities.

The framework for constructing and maintaining situated spatio-temporal representations and concepts is introduced into the overall system of WP7. From an integration point of view, hybrid map layers need to be constantly modified with new data being acquired or new place and concept descriptors being inferred. In the scope of the reasoning process, the multi-layered map repository is required to facilitate efficient information retrieval.

Partner	Contributions
ETHZ	hybrid maps, spatial and spatio-temporal representation layers
DFKI	domain inference over spatio-temporal representation layers for establishing topological place graphs and conceptual descriptors
CTU	localization using omni-directional vision
BLUE	localization using odometric information
Fraunhofer	localization using range sensor data, geometric models for spatio-temporal representations
Rome	temporal reasoning, symbolic anchoring of time-space constraints and compositional structuring of functional representation.

Description of work – Risk management

Probability: Low (unlikely to occur); medium (not unlikely to occur); high (likely to occur)

Gravity: Low (uses available mechanisms within system); medium (uses available technology, not yet in system); high (requires new technology).

Risk	Probability	Gravity	Contingency Plan
The environment may lack structure to apply semantic concepts to.	medium	low-medium	Use clarification strategies and running commentary to make clear what the robot understands; use targeted help in establishing references to the environment [90]
Multi-modal and distributed sensor sources, may require dealing with data inconsistencies and incoherences at a new level of complexity.	medium	low-medium	We will model sensor characteristics and apply probabilistic sensor fusion techniques whilst injecting suitable world assumptions. Also, we expect sensors to further improve during the project term.
The spatial modeling plays a central role within the scope of the overall project, with all other WPs relying on WP1 results.	medium	high	We may fall back to tentative on-line data processing strategies based on simulated environments or provide offline results from real-world observations.
High data processing demands and large data amounts may limit implemented model response times.	medium	medium	The concept of shared control leaves some margin for timing requirements. Efficient data processing may thus focus on time-critical model aspects. Use parallelization.

Description of work – Evaluation

The need for consistent and coherent spatio-temporal models poses a central challenge to the research performed within WP1. Evaluation of these models becomes an essential part of our efforts. Unfortunately, there is no straightforward universal solution to the problem of assessing spatio-temporal models with respect to the real world. This is usually due to a lack of ground truth information which is difficult to acquire. In WP1, we aim for a threefold evaluation process.

Quantitative map consistency evaluation:
For the spatial and spatio-temporal layers of the map hierarchy, we validate our modeling strategies against simulated ground truth data (see e.g. [123]). WP1 performs computational consistency checks and determine quantitative accuracy measures. Under dynamic world assumptions where accurate measurements become exceptionally difficult, simulations will complement field studies. In both conditions, the metrics will include repeatability and precision. The repeatability can be tested by different trials in the same conditions, whereas precision will be checked against some given reference points. The typical deviation we aim at are below 10 cm.

Qualitative map consistency evaluation:
In contrast, topological and conceptual layers of the map hierarchy are assessed in a qualitative manner, by validation against manually-labeled real-world results. A more detailed metric is the recognition rate of a place. In simulation, we aim at values in the range 80%-95%. In a real environment, we aim at a range of 65%-90% to cope with the increased complexity.

Human-level map evaluation:
Finally, in accordance with the overall objectives of NIFTi, we want to validate the development process in an iterative manner. Formative evaluations will be lead yearly with end-users and their feedback will be addressed in the subsequent development processes.

Deliverables:

The software prototypes developed in WP1 serve directly as input to WP7.

DR 1.1.1 *Acquisition of spatial maps of semi-structured environments.* R. (M12) [36.25 PM]

DR 1.2.2 *Acquisition of spatio-temporal maps and place topologies of semi-structured environments*. R. (M24) [24 PM]

DR 1.3.3 *Acquisition of human-compatible hybrid maps of semi-structured environments*. R. (M36) [27 PM]

DR 1.4.4 *Spatio-temporally grounded situation awareness using a-priori information*. R. (M48) [21 PM]

1.3.5.2 WP2: Visuo-conceptual modeling for situation awareness .

Work package number:	2	Starting date or starting event						1
Work package title	Visuo-conceptual modeling for situation awareness							
Activity type	RTD							
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	12	0	16	0	8	55	9	0	0

Objectives

WP2 contributes to achieving Objective 1. It provides observations about static and dynamic aspects of the robot's surroundings, focusing on landmarks, terrain features, threads, victims. The WP2 objectives are:

(A) Gather and implement state-of-the-art computer vision modules needed for basic vision functionalities for getting percepts from a perspective camera, a stereo camera pair on a pan-tilt unit and from an omni-directional camera.

(B) Develop image/video understanding capabilities allowing to ground raw percepts to higher level conceptual representations.

(C) Develop methods for adaptive detection of objects/landmarks in the scene both in static images and video.

(D) Develop methods for analyzing percepts from a stereo camera head, and controlling it.

(E) Develop methods for understanding omni-directional camera percepts for robot navigation and new event detection.

(F) Develop methods for detecting and tracking of humans, victims in video.

The case of the moving observer will be of our interest as well. The basic idea (used in our previous projects) is to segment objects and register their background in the input sequence. After background registration, frames can be treated as if taken by the static camera.

Description of work – Tasks

Task T2.1: *General computer vision and audio functionalities [25 PM]*
(Period: M0-M48, Partners: [CTU/ 9 PM, Frauenhofer/ 6 PM, ETHZ/ 6 PM, Roma/ 2 PM], Contributes to objectives: A)
Initially, readily available background knowledge/implementations, open source and off-the-shelf tools are used. Methods cover image segmentation, object recognition in images, motion detection (both from a static camera and a moving one) [230]. The challenging cognitive architecture and robot mission in an unknown semi-structured environment call for development of new methods too. This part constitutes the scientific contribution of the task. The robot has an open audio stream on-board. WP2 uses ready solutions for detecting audio events as crash or a victim crying 'help'. These audio events are synchronized with vision and transferred to WP1.

Task T2.2: *Image/video understanding capabilities and symbol grounding [18 PM]*
(Period: M3-M42, Partners: [CTU/ 10 PM, DFKI/ 6 PM, ETHZ/ 2 PM], Contributes to objectives: B)
Symbol grounding provides meaning to percepts and contributes to creating conceptual representations, which can be used in symbolic reasoning. WP2 extends approaches which CTU and DFKI have developed, in conjunction with the concept-oriented methods for grounding and interconnecting content as developed in WP3. WP2 involves categorization efforts, and explore

intermediate representation levels as suggested in [92]. This allows using two ontologies simultaneously: a visual concept ontology and an image processing ontology. Learning plays important role here. NIFTi builds on top of CTU's efforts combining statistical and structural pattern recognition approaches [220], performing image segmentation jointly with image interpretation exploring structure [205]. NIFTi starts from the recently proposed method to grounding in image interpretation which has been formulated as an optimization task [64].

Task T2.3: *Adaptive object/landmark detection in images/video [20 PM]*
(Period: M3-M42, Partners: [CTU/ 8 PM, Fraunhofer/ 10 PM, ROMA/ 2 PM], Contributes to objectives: C)
View-based object detection based on local features has been an important paradigm change in vision [45, 159]. The approach based on detection of MSER (maximally stable extremal regions) [153] has very good properties. The idea has been generalized in CTU to videos [155, 281]. Based on this idea, NIFTi develops an on-line learnable object tracker/detector.

Task T2.4: *Stereo camera head, image understanding and control [14 PM]*
(Period: M6-M48, Partners: [CTU/ 10 PM, DFKI/ 4 PM], Contributes to objectives: D)
The stereo head is mounted on a pan-tilt unit. The head allows the robot active visual exploration. T2.4 develops methods needed for analysing still images/videos and controling the robot head. T2.4 also contributes to terrain classification, e.g, to detect a clear corridor pathway ahead, a floor surface with spotted debris which cannot be passed, or a hole in the floor or wall.

Task T2.5: *Omni-directional camera for navigation and event detection [10 PM]*
(Period: M6-M48, Partners: [CTU/ 10 PM], Contributes to objectives: E)
NIFTi uses an omni-directional camera with a 360° viewing angle. T2.5 uses the zero-phase of the frequency representation for the omni-directional vision as a magnetic compass [194]. This method gives the possibility to orient in the environment. T2.5 applies the sensor for the NIFTi needs and develop methods for visual SLAM and detection of unexpected events (including threats) in the robot surroundings. The latter will be used by the WP4 attention mechanisms.

Task T2.6: *Human detection/tracking in video [13 PM]*
(Period: M3-M48, Partners: [CTU/ 10 PM, ROMA/ 3 PM], Contributes to objectives: F)
T2.6 first involves CTU's existing implementations which comprise motion detection methods for static camera, moving human detection and model-based tracking [124]. T2.6 also provides modules for detecting human victims [247] in the scene, detecting both conscious and unconscious people.

Description of work – Milestones

Milestone MS2.1: *Basic functionalities of the computer vision subsystem* (M10)
Basic computer vision functionalities are implemented into the NIFTi integration system. The implemented method consists of basic functionalities from tasks T2.1-T2.6 focusing on those which were available on project start. The interfaces with other WPs are established and tested. Methods needed for the Yr1 integrated system are integrated with the stress on landmark detection and localization (WP7).

Milestone MS2.2: *Vision for human assisted exploration* (M22)
Vision methods needed for 3D descriptions of the robot environment are provided. This includes an operational stereo vision head on a pan-tilt unit and algorithms for stereo data processing and robot head control. The omni-camera and the gathering of information from it is made operational. The link to higher levels via percepts grounded to symbols is enabled and its basic functionalities are provided. The feedback link from higher level WPs is established. The ob-

ject/landmark detection methods are enhanced. Basic human detection and tracking capabilities are integrated into the system. Methods needed for the Yr2 system are integrated (WP7).

Milestone MS2.3: *Vision for shared situation awareness* (M34)
All tasks deliver functionalities allowing to deal with more complex rescue scenarios and share knowledge across the system. The core issues are the interplay between symbol grounding and feedback provided to WP2 from higher level knowledge management system. The capabilities in adaptive object/landmark detection, stereo vision head information, omni-based navigation, motion analysis with the stress to human detection and tracking are demonstrated in mature form. Methods needed for the Yr3 system are integrated (WP7).

Milestone MS2.4: *Vision for the final prototype* (M44)
The efforts in all tasks will be concentrated towards the best performance of the final prototype. Functionality of perception-action loops and scaling of the solutions to more difficult situation will be demonstrated.

Description of work – Contributions to project & state-of-the-art

The role of WP2 is to obtain task relevant information from visual percepts, taking into account the current context, and provide this information to other WPs. The recipient of this information is mainly WP1, which serves as a bidirectional interface and a master for WP2. The relation to other WPs can be seen from the figure in the description of WP1. Notably, the work on grounding of visual conceptual descriptions contributes to the conceptual layer in that figure, and interacts with the mechanisms developed in WP3 for interconnecting conceptual content. There is an important link with WP4 which provides selectional attention mechanisms, priming the processes developed in WP2 where to look, and what to look for.

Partner	Contributions to WP2 efforts
CTU	visual percepts from ordinary cameras, stereo camaera head and omni-directional sensor capturing video or still images on demand; trainable object detector to find landmarks, detection of humans in video, creating symbols from percepts
Fraunhofer	real-time visual recognition, visual classification of terrain features
BLUE	sensor fusion
DFKI	symbol grounding
ETHZ	grounding of observation models in spatial models
ROMA	multiple viewpoint recognition, victim recognition

Description of work – Risk management for WP2

Risk	Probability	Gravity	Contingency Plan
Compatibility problems of background knowledge modules delivered by diverse partners.	low	low	Some of the modules are to be re-implemented or re-interfaced.
Vision problems to be solved in the real rescue scene are too challenging.	high level	medium	This always happens in challenging settings like NIFTi. Progress towards full functionality is gradual. Meanwhile, simplifications in the scene are explored. After many trials, risk assessment based on probability analysis will be used.
Symbol grounding task is too challenging to deliver representations useful in higher level reasoning.	high level	medium	Close cooperation with higher level modules is needed to specialise the image understanding methods.
Detecting threats and victims are difficult: too many unpredictable ways to appear.	high level	medium	WP2 concentrates on typical cases and gradually extends threat and victim detecting capabilities.

Description of Work – Evaluation

General comments on evaluation of vision modules
Vision modules can be validated at each moment of the project by verifying the performance on demonstrators. Before developing the method its validation is always prepared. This development validation cycle is used incrementally. Quantitative measures of success will be used when possible, e.g., false positives, false negatives, ROC curves, stability versus various data errors and noise. There will be also need for evaluation against human expectations and abilities of a human operator. We will have to do field tests with end users and evaluate results of our machine versus human both qualitatively and quantitatively. We have to quantify how and to what degree can the automatic analysis be used to increase the situation awareness of the human.

Validation of vision modules and their inter-connectivity to other modules:
The particular vision functionality is tested with regards to the module which uses the results.

Landmark-based omni-directional-based vision navigation
The performance is compared to established laser-based navigation and expressed quantitatively in false positives/negatives..

Threads and victim detection:
The evaluation is conducted on real videos taken from training disaster sites.

Deliverables:

The software prototypes developed in WP2 serve directly as input to WP7.

DR 2.1.1 *Basic vision-based situation awareness capabilities*. R. (M12) [23PM]

DR 2.2.2 *Stereo- and omni-directional vision for human assisted exploration*. R. (M24) [23PM]

DR 2.3.3 *Bi-directional cooperation of low-level vision modules and higher level control*. R. (M36) [25PM]

DR 2.4.4. *Scaling the functionalities of vision subsystem with the complexity of the environment*. R. (M48) [29PM]

1.3.5.3 WP3: Adaptive multi-modal HRI for joint exploration .

Work package number:	3	Starting date or starting event							1
Work package title	Adaptive multi-modal HRI and flexible planning for joint exploration								
Activity type	RTD								
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	50	12	9	0	0	0	19	0.75	1.5

Objectives

WP3 provides the functionality for a cognitive architecture to communicate with a human in a robust, adaptive fashion. With that, WP3 contributes to Objective 3. Communication is multi-modal, using spoken dialogue and a GUI (visual display, gesture). A NIFTi cognitive architecture uses communication for joint planning and re-planning, and keeping up a running commentary. The commentary provides environment descriptions, plan progress, anticipatory warnings, and means for clarifying situation awareness. Communication is grounded in several ways. Referential descriptions are resolved against functional environment models and the situated exploration history; plan descriptions are interconnected with plans; variations in communicated attentional focus are linked with perceptual attention; and variation in communication style and associated linguistic cognitive load are connected with cognitive user models. With this, WP3 addresses the human factor in communication: how the scope, content, and form of communication (dialogue, GUI presentation, fusion across modalities) can be adapted to anticipated cognitive task load and attention given the circumstances (plan, environment). The WP3 tasks address three objectives:

(A) Robust spoken dialogue comprehension for multi-modal HRI.

(B) Production of adaptive multi-modal HRI for joint exploration.

(C) Strategies for clarification and adaptive grounding.

Description of work – Tasks

Task T3.1: *Robust spoken dialogue comprehension for multi-modal HRI [32 PM]*
(Period: M0-M48, Partners: [DFKI /20 PM,TNO/9 PM, FDDo/0.75PM, VVFF/1.5PM], Contributes to objectives: A)

Task T3.1 focuses on robustly comprehending multi-modal HRI. It develops the means for using context to guide processing spoken language, and to provide indications about cognitive task load. T3.1 uses domain studies (DR3.1.1, DR3.2.3) to build up corpora of domain-specific human-robot interaction (Wizard-of-Oz). Spoken interaction is transcribed, and annotated with features modeling grammatical information (syntactic, semantic interpretation), dialogue-level interpretation (speech act, referential links), and current task load (computed using the CTL models of WP4). Grammatical information is encoded using Combinatory Categorial Grammar (CCG) [6, 5] (DFKI). The corpora are used to train semantic language models for automatic speech recognition, and discriminative models for parse selection [147, 148] (DFKI).

T3.1 develops a novel approach to integrating speech recognition, parsing, and dialogue interpretation. The approach incrementally processes spoken dialogue for HRI [25], [148, 131] (DFKI). The point of seeing processing of spoken dialogue as an integrated process makes it possible to better deal with typical problems in spoken dialogue such as recognition quality, ungrammatical or incomplete utterances, and pervasive ambiguity. Speech recognition is a hard problem; rather than trying to solve it in isolation, NIFTi combines speech recognition with other levels of processing to help alleviate problems, and improve overall performance. [147] show how speech recognition can be improved by making it context-sensitive. Using information about salient objects, events, and places in the current context, lexical items are activated. These lexical activations in turn balance the language model to "listen to what is most likely to be heard" (while still allowing for backing off to less likely interpretations). [148] propose to use the resulting word lattice directly in a context-sensitive, incremental parsing process. This parsing process combines a linguistic grammar with (a) non-standard rules for combining expressions, and (b) discriminative models for parse selection [280]. The non-standard rules extend coverage to deal with missing words, dysfluencies, discourse-level composition, and correcting typical speech recognition errors. Parse selection models help to select the probabilistically most likely analyses, given the linguistic structure, content and context of the utterance. The parse selection models are learnt using a perceptron model trained on a domain corpus, and are applied after each incremental parsing step. In NIFTi, these initial developments will be developed further. To deal with stress-induced changes in voicing NIFTi intends to use stress detection for adaptive speech recognition [82, 197, 104]. To address time-criticality, NIFTi considers the combination of the symbolic parser with statistical parsers for CCG [38] to achieve parsing speeds for regular utterances of less than 200ms, and the use of a statistical control to guide processing for speech recognition, parsing, and dialogue interpretation. Its purpose is to establish those points during incremental processing, at which the combination of information from different sources can be optimally used to disambiguate, rank, and complement the current set of analyses within a process. This avoids the architecture spending unnecessary resources to little effect. Information sources include these levels of interpretation, and information about how content (established so far) can be resolved and grounded in the larger context ([100, 131, 148] (DFKI), T3.3). Controller policies are learnt off-line on corpora, using reinforcement learning [87] and statistical relational learning [68].

From year 3 onwards, T3.1 includes vision-based gesture recognition for human-robot interaction. The vision technology is based on the functionality developed in T2.6 (§1.3.5.2). The purpose of the gesture recognition is twofold: One, to allow for hand-signal communication as in [119], and two, to provide for basic gestures to indicate areas or directions. In the architecture a direct channel will be used to provide information about visually recognized and classified gestures to the working memory of the HRI subarchitecture (cf. §1.2.6). Here, this information is combined with other (timed) information (e.g. speech) to interpret the communicative intention of the gesture in context.

T3.1 further increases robustness by adapting speech recognition to stress factors. T3.1 integrates audio cues regarding stress levels [82, 197], with models for relating speech and cognitive task load [104] (DFKI), and the semantic language models trained on corpora with task load as specific feature. The resulting predictions about perceived stress level are provided back to the cognitive user models (WP4).

Task T3.2: *Adaptive multi-modal HRI for joint exploration [30 PM]*
(Period: M0-M48, Partners: [DFKI/12 PM,TNO/3 PM,Fraunhofer/9 PM,ROMA/6 PM], Contributes to objectives: B)

Task T3.2 develops means for producing adaptive, multi-modal HRI. The goal of adapting content presentation is to optimally align with a user's cognitive task load and attention. Adaptation regards how the cognitive architecture produces communication – including verbal communication, visual presentation, and the distribution of content presentation over these modalities.

T3.2 integrates the production of verbal communication into the broader task of models of collaborative dialogue to cover the joint exploration domain [149, 80, 18, 79]. These models, together with how content can be grounded over space and time (T3.3, WP5) serves as "common ground" in the running commentary. T3.2 uses flexible methods for planning dialogue [24, 23, 22], and connects the planning domain with work flow models (WP4). Production then combines decision models for selecting the appropriate context for an utterance [195], [278](DFKI) (given the ongoing collaboration), what content should be included [41, 129], and how this content should be linguistically realized [268]. Content in T3.2 focuses primarily on verbalizing plan progress and situation awareness for the running commentary [116, 278]. In the decision processes for production (modeled as decision trees), T3.2 integrates linguistic constraints [129], models of linguistic processing load [126], [116](DFKI), and input from cognitive user models on cognitive task load and attention (WP4). Verbal production thus balances possible linguistic variation (modeled in CCG [6, 5]) with methods for choosing an optimal variant in a given context. Optimality means minimizing impact on cognitive task load while still being maximally informative. The decision models can be trained, using reinforcement learning [212, 141]. T3.2 uses domain studies to gather training data (Wizard-Of-Oz setup; DR3.1.1, DR3.2.3) for offline training, and investigates how training can be extended to include online data.

T3.2 combines adaptive verbal production with adaptive multi-modal GUIs [103]. The purpose of the GUIs (in-field PDA or remote laptop) is to provide a rescue worker with visual information about the cognitive architecture's situation awareness and the joint plan [60, 59]. GUI design follows the insights from [29, 72, 74]. The rescue worker can use the GUI to share control with the cognitive architecture, using ecological interfaces [186, 73], and pen-based gesture to aspects of the joint plan and the visuo-spatial information [108, 109]. As in verbal production, adaptivity regards the selection of the appropriate context (situated context and planning context), content selection, and presentation (e.g. which part of a plan, or for a map: scale, points of interest, focus) [210, 4, 260, 259]. Because visual presentation and verbal production can accompany one another, T3.2 considers how to optimally balance distribution and repetition of content over different modalities ("fusion") given perceived cognitive task load and attention in an integrated fashion [210, 102] (DFKI), [213].

Task T3.3: *Clarification and adaptive grounding [31 PM]*
(Period: M0-M48, Partners: [DFKI/18 PM,ROMA/13 PM], Contributes to objectives: C)

Task T3.3 focuses on grounding content from multi-modal interaction, planning and cognitive user models in functional environment models. Grounding involves resolving and interconnecting content, and clarification. Resolution establishes how plans and communicated content refer to the environment. Interconnecting establishes relations between content across these modalities. Relations are maintained over space and time, so changes can be monitored. Clarification is used to resolve uncertainty or incompleteness in establishing how to ground content.

T3.3 develops a new approach for adaptive grounding. The approach is based on a combination of a Bayesian, graph model-based approach to modeling information fusion, and learnable controllers. The Bayesian approach is a probabilistic reformulation of DFKI's [100]. [100] based grounding on the idea that subarchitectures write a-modal representations of content (called proxies) to a working memory, after which these proxies get bound (grounded) into unions. Binding was done on a symbolic feature-by-feature comparison. The Bayesian approach reformulates proxies as content structures with associated multivariate probability distributions. These distributions indicate possible hypotheses for how features of that content could be interpreted. This adequately represents the uncertainty in interpretation. Furthermore, the approach provides the possibility for features across any number of proxies to be compared, and to use both discrete and continuous features. Given proxies, features and probability distributions, the probability of a union given the evidence of the modal observations of the proxies is then computed by conceiving of the feature distributions as an independent likelihood pool. The underlying prior probabilities are computed using a multi-modal directed graphical model, i.e. a Bayesian network of feature probabilities. The role of this network is to specify the dependencies or correlations between feature instances. Furthermore, they offer a strong theoretical foundation for a robust and adaptive approach to symbol grounding, with several efficient machine learning algorithms in existence for learning both the parameters and the structure of such models. Their graphical nature provides a stronger model for grounding than aforementioned mediation models, as it can in principle capture dependencies over structures (not just individual nodes).

Learnable controllers are used to adapt three different aspects in grounding. One, controllers are used for online learning of the parametrisation and structure adaptation of the graph models. This establishes a basic a-modal (conceptual) level of content description that fuses information across different modalities. A second use is to learn how changes in fused content can be tracked across space and time. Finally, controllers are used to learn how different content descriptions can be resolved to identical aspects (referencing, aliasing). Each of these types of controllers establishes bi-directional associations (cf. [234, 233]) between the grounding levels in the environment model (WPs 1,2) and complex conceptual structures. Associations allow for multiple complex mappings, to enable aliasing. Changes in conceptual content (agency) and in a dynamic environment can be percolated through the controllers to update grounding status over time. The development in how content and environment model have been associated is maintained in the situated exploration history (WP5: T5.5). Adaptation combines online learning mechanisms from a multi-level approach to grounding [234, 233, 100] and reinforcement learning [271](DFKI) to deal with low data volumes. Adaptive grounding covers the grounding of plans and anticipated states [23] (DFKI), visual aspects of the environment and their functionality [40, 214, 215], spatial and spatio-temporal referential descriptions [115, 116] (DFKI). Grounding is sensitive to perspectivization [250].

Content is represented uniformly in a logical formalism. The representation includes temporal primitives for modeling temporal structure and temporal reasoning [130, 128] (DFKI), and statistical primitives to manage approximate inference and uncertainty for initial knowledge [202, 162], and action execution [217]. Domain inferences over these representations is used to expand content with associated concepts [44] and [133] (DFKI).

T3.3 extends adaptive grounding with explicit methods for clarification. Clarification is a dialogue-based mechanism for resolving uncertain or incomplete information – whether from communication, perception, or planning. T3.3 incorporates clarification strategies in dialogue planning (T3.2), inspired by [251, 137, 206, 229], [132, 70] (DFKI). T3.3 explores learning methods for adapting clarification strategies online [211, 271, 212] to align to changes in describing and referring to a novel environment.

Description of work – Milestones

The milestones define measurable capabilities to be delivered by this WP. (See below for evaluation metrics.) These capabilities incorporate the functionality provided by the WP tasks, and contribute directly to the milestones for the integrated systems of the roadmap (§1.1.5, §1.3.7).

Milestone MS3.1: *Dialogue-based HRI for human-instructed navigation* (M10)
The cognitive architecture can understand spoken dialogue for movement commands and route descriptions for human-instructed navigation (remote). Dialogue includes references to landmarks, static threats, and spatial structure. Reference resolution is based on a domain analysis of how rescue workers refer, to establish reference points in disaster environments. The GUI (laptop) visualizes the cognitive architecture's hybrid maps, and provides pen-gesture to spatial topology, landmarks and terrain features. Shared control uses way-point navigation. Adaptation focuses on grounding spatial references, and parameterization of movement actions to adapt to new terrain. The system can clarify navigation alternatives and spatial references, and produces a commentary of what it sees and does.

Milestone MS3.2: *Multi-modal HRI for human-assisted navigation* (M22)
The cognitive architecture has extended dialogue capabilities to cover joint plan construction and negotiation for exploration planning (remote). Dialogue includes spatio-temporal reference to static and dynamic aspects of the environment. Reference resolution is based on a further domain

analysis. Adaptation mechanisms for grounding are extended to ground plans, anticipated states, and dynamic spatial aspects. The GUI visualizes the plan and its monitored execution. The running commentary is extended to deal with online re-planning.

Milestone MS3.3: *Multi-modal HRI for in-field joint exploration* (M34)
The cognitive architecture interacts with an in-field rescuer, who coordinates the planning of jointly exploring the environment. Dialogue includes production, resolution of spatial- and spatiotemporal references under agent- and object-centric perspectives, focusing on joint planning. The GUI is ported to a portable device (PDA). The cognitive architecture keeps up a running commentary of what it sees and does, from its own perspective. Clarification and adaptive grounding includes grounding perspectivized descriptions and actions, and terrain-relative affordances for navigation.

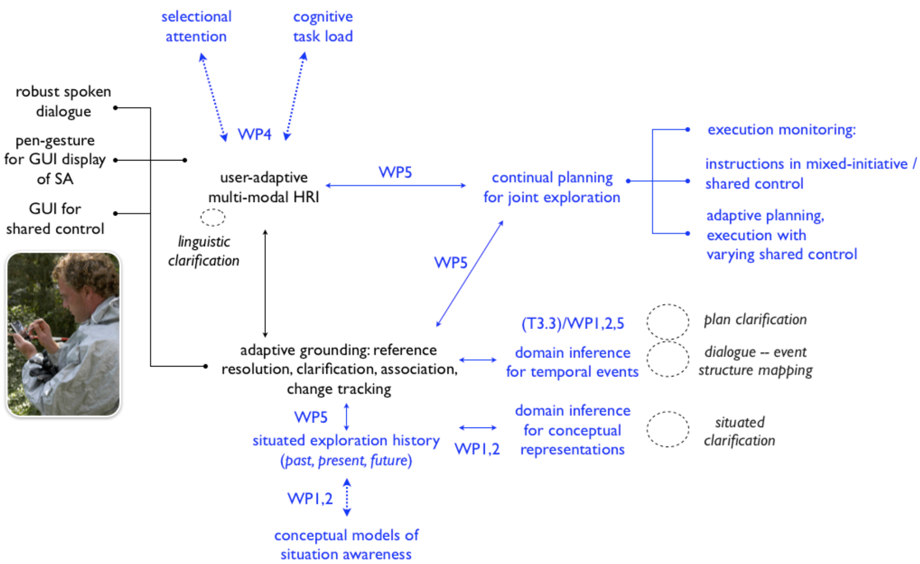
Milestone MS3.4: *Multi-modal HRI for sharing situation awareness* (M44)
The cognitive architecture interacts with in-field rescuers, to jointly plan exploration and to share situation awareness. Reference production and resolution are extended to include perspectivization for sharing awareness of spatiotemporal aspects of the environment. The cognitive architecture is able to process a rescuer's commentary, to update the joint plan and its own situation awareness. Adaptive grounding is extended to handle perspectivized descriptions and actions.

Description of work – Contributions to project & state-of-the-art

T3.1 contributes a novel method for robustly processing spoken dialogue in HRI. It moves beyond the state-of-the-art in speech processing and parsing of spoken dialogue by considering the integration of the these processes together with dialogue interpretation, and to dynamically control information and processing at these different levels. Control policies are learnt off-line from domain corpora gathered in this WP, and are adapted online using reinforcement learning and statistical relational learning. T3.1 advances the integration of cognitive user models into cognitive architectures by developing a bi-directional link between such models and stress detection for adaptive speech recognition. T3.2 takes this further, developing a close coupling between user adaptation, and producing multi-modal communication for human-robot interaction. T3.2 addresses the unique challenge of adapting communication in a highly dynamic setting. All agents are dynamic, communicating over a longer period of time, and operating in a dynamic and unpredictable environment. This moves T3.2 well beyond the desktop metaphor. T3.2 shows how adaptation and alignment evolve over time, varying with the user's cognitive make-up under influence of operating in a dynamic environment. T3.1 and T3.2 move multi-modal HRI well beyond the current state-of-the-art through a combination of advanced, user-adaptive spoken dialogue and multi-modal GUIs. Together with the situated exploration history (WP5) this provides HRI the possibility of communicating about long-term temporal perspectives.

T3.3 further advances NIFTi's approach to "symbol grounding" (together with WPs 1,2). It moves beyond grounding of single categories. It advances on categorical complexity, grounding entire conceptual structures and not just individual concepts. Furthermore, it grounds concepts in space and time, and can track changes over space and time (in conjunction with the situated exploration history). ("T3.3 moves beyond just ventral to include dorsal.") Finally, T3.3 includes clarification as an explicit mechanism in grounding, to resolve uncertainty and incompleteness.

The next figure shows how the main functionality in WP3 is integrated into the cognitive architecture, and connections to other WPs.



Partner	**Contributions to WP3 efforts**
DFKI	situated spoken dialogue processing for HRI, multi-modal dialogue, intelligent user interfaces, user adaptivity, adaptive mechanisms for reference resolution and grounding, temporal reasoning
TNO	interface design for USAR, intelligent user interface, multi-modal dialogue, user adaptivity
ROMA	representations for adaptive grounding for planning and execution monitoring, linking dialogue and planning
Fraunhofer	intelligent user interfaces

Description of work – Risk management for WP3

Probability: Low (unlikely to occur); medium (not unlikely to occur); high (likely to occur)

Gravity: Low (uses available mechanisms within system); medium (uses available technology, not yet in system); high (requires new technology).

Risk	**Probability**	**Gravity**	**Contingency Plan**
Difficulty in speech recognition	medium-high	medium	Improve offline/on-line speaker training; let cognitive architecture instruct human how to communicate with it [90]; complement symbolic parsing with statistical parsers
Difficulty in resolving dialogue- and planning referents	medium	low-medium	Clarification: request use of GUI pen-based gesture to provide further information about references to environment, and adapt grounding
Possibility of non optimal planned strategies, in presence of unidentified events/situations	medium	low-medium	Use dialogue to augment situation awareness. Context based reasoning to infer missing information

Description of work – Evaluation metrics

To evaluate progress on the tasks, WP3 measures improvements on robustness in recognition and interpretation, and on task efficiency. Because USAR is a new domain for multi-modal HRI including situated dialogue and planning, there are currently no testbeds or benchmarks against which to compare. WP3 creates corpora and test beds, and makes them available for public use.

Robust processing in HRI

Word-error rate (WER) for speech recognition: using context to determine optimal language

model predicts lower WER over using single language model. *Precision/recall* for interpretation: using context to integrate partial analyses in utterance understanding predicts a higher number of integrated analyses (recall) reflecting the intended interpretation (precision). P/R can be measured from test beds with prototypical use case interactions. Indirectly: more robust processing predicts less need for dialogue clarification, resulting in a higher task-efficiency.
The target WER ranges from 25% down to 10%. Context-sensitive ASR [147, 148] currently achieve a WER of about 15%, on free spoken dialogue on a limited domain. Where the actual WER is within this bandwidth depends on the complexity of the dialogue: For years 1 and 2 we aim for a target WER in the range of 18-10%, for years 3 and 4 25%-15%.
More important than WER in isolation is the eventual construction of the intended interpretation. [148] currently achieve 67% in total match and about 87% in partial match (both nbest 5 F_1-measure), which establishes a state-of-the-art baseline. NIFTi particularly needs high-precision interpretations, correctness being more important than getting just some interpretation. Target precision scores on partial match are therefore in the range of 90-95%, with target F_1-scores 70% (exact match) and 90-92% (partial match).

Efficiency of HRI

Task-efficiency: adaptively grounding dialogue in the user-, situated- and task-context predicts a higher task-efficiency in dialogue, as the cognitive architecture understands what actions and objects are talked about. Deterioration in grounding leads to a higher need for clarification, resulting in a decrease in task efficiency. Task efficiency is measurable with the PARADISE evaluation scheme for dialogue systems [266], in user experiments. The efficiency of clarification can also be measured in the number of turns needed to resolve the clarification (PARADISE). The better the architecture is able to formulate a clarification request, the fewer turns are needed.
Efficiency in grounding particularly relies on being able to resolve (and produce) spatiotemporal references. The reference resolution and production tasks become progressively harder throughout the project. This is related to the increase in terrain complexity, and the increased difficulty for the robot in perceiving the environment (building up maps, observation models). Target partial match for the linguistic construction of the intended semantics of a human's referring expression is 85-95%. NIFTi targets resolution and grounding of such referential semantics at 65-90%. These targets are provided with the provisos that (a) these targets are dependent on the quality of the underlying perceptual models, and (b) achieved performance may vary by type of referring expression and context.
The target for clarification efficiency can be measured on the number of follow-up turns required for reformulating or extending a clarification request. The target here is to achieve maximally two (2) human follow-up turns.

Effectiveness of HRI

Task-effectiveness: How accurate and complete is the task performed. An higher effectiveness is reached when cognitive architecture and human understand each other completely accurately. The aim here is to minimize the time robot and human need to spend on meta-communication – trying to understand what each is talking about. The target here is to spend maximally 10-15% of total operation time on meta-communication.

Satisfaction of HRI

Task-satisfaction: People can be satisfied with a product even when it does not have the best performance. This depends for a large part on how errors are handled and what they initially expected of the product. Satisfaction will be measured using standard questionnares (e.g. [151]).

Deliverables:

In years 1 and 2, domain analyses are performed. They describe field experiments for gathering dialogue data, and their analysis. Results are included in reports DR3.1.1 and DR3.2.2 (listed below). The software prototypes developed in WP3 serve directly as input to WP7.

DR 3.1.1 *Adaptive situated HRI for human-instructed navigation.* R. (M12) [23.25PM]

DR 3.2.2 *Adaptive situated HRI for human-assisted navigation.* R. (M24) [23 PM]

DR 3.3.3 *Adaptive situated HRI for in-field exploration planning.* R. (M36) [23 PM]

DR 3.4.4 *Adaptive situated HRI for in-field joint exploration.* R. (M48) [23 PM]

1.3.5.4 WP4: Cognitive taskload and selectional attention .

Work package number:	4	Starting date or starting event							1
Work package title	Task load and selectional attention								
Activity type	RTD								
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	20	55	3	0	0	4	8	5.5	6

Objectives

The objectives of WP4 are to develop:

(A) Cognitive taskload-based mechanisms for scheduling and allocation of tasks.

(B) Mechanisms for selectional attention based on mission briefs (presence of people, dangerous items) and context.

(C) Mechanisms for determining the right time and right form of multi-modal dialogue based on situated context, user models (workflow), and cognitive task load.

These contribute to Objectives 3 and 4.

Description of work – Tasks

Task T4.1: *Develop mechanisms for scheduling and allocation of tasks based on task load [32 PM]*
(Period: M0-M48, Partners: [TNO/20 PM,DFKI/3 PM,ROMA/3 PM, FDDo/2.5PM, VVFF/3PM], Contributes to objectives: A)

Task 4.1 develops mechanisms for joint planning and allocation of tasks based on the (morphological) capabilities and availability of user and robot. With tasks that can be considered for both robot and human T4.1 makes it dependent on the user's cognitive task load. Whereas T5.1 aims at an efficient deployment of resources over time (scheduling resources for operation), T4.1 focuses on a transparent adaptation of the tasks-to-do towards the human cognitive characteristics (task harmonization). The cognitive taskload should be neither too low nor too high. Otherwise people can reach for instance cognitive lock up, meaning they only pay attention to the task they are currently working on and do not switch to higher-priority tasks by themselves [175, 49, 50].

The subsequent step is to support dynamic task allocation for harmonizing user's momentary cognitive taskload to the current situational demands. A guiding principle is that humans make working agreements for the adaptation during the preparation and planning of a mission (e.g., under which conditions the level of automation changes; cf. [179]). The adaptation should be further refined by attuning the scheduling of tasks to the contextual dynamics and possible suboptimal human states (e.g., fatigue). Eventually, T4.1 extends further cognitive load balancing in HRI for shared control (i.e. [58, 28]) to joint planning and exploration using mechanisms for providing the right information at the right time and in the right form (Task 4.3).

We use a task- and user models for adaptive HCI as a basis [78]. The task model includes, among other things, a description of cognitive task load an their effects on performance, mental effort and emotional state [175, 178]. Task demands and emotional states are described at three levels. The first level specifies the human act observables that correlate with human information processes (HIP). At the second level, HIP dimensions represent variables that correlate with human performance. For cognitive task load, the dimensions are percentage time occupied, the level of information processing [208], and the number of task-set switches. SOWAT, an activity monitoring tool, is used to derive the CTL-dimensions values from observables as user-interface acts [78], while affective computing techniques are used to derive the emotional-state from, for example, speech expressions [256]. The user model is applied for personalized estimation of HIP-dimensions values from observables. At the third level, HIP classes are derived from the dimensional models. For example, CTL classes are underload, overload, vigilance, cognitive lock-up, and neutral; emotional-state classes are boredom, relaxed, excited, stressed, and neutral.

The task- and user model for adaptive HCI that we use has been applied in a demanding dynamic indoor environment [76], but never in an outdoor environment which puts extra demands on the models. Task switches, for instance, are harder to recognize in the USAR environment. To estimate which tasks are active, possible solutions are to combine the operation plan (including the briefing results), the location and context information, the communication patterns, and the direction the human worker moves [237].

Task T4.2: *Develop mechanisms for selectional attention using current context and knowledge [22 PM]*
(Period: M0-M48, Partners: [TNO/5 PM,DFKI/5 PM,Fraunhofer/3 PM,CTU/4 PM,ROMA/5 PM], Contributes to objectives: B)

A-priori knowledge, knowledge acquired during deployment, and the current context all influence where a robot should look. A robot which knows it is on an earthquake site looks in cavities for survivors, because the chance of survivors is high there. Furthermore, a robot could have its attention directed by the user, by GUI or dialogue. This helps in reducing the feature space dimensionality and combining both temporal and spatial aspects, (extending approaches dealing mostly with static scenes, e.g.[99, 185, 257, 275, 253]). T4.2 generates a saliency map both on the basis of a search model and on the "region pointing" suggested and solicited by the human operators [13, 14, 11, 31]. This aims to decrease the human cognitive taskload because the "pointing" based interaction does not require the human to compel his/her visual search but only to indicate to the robot where to orient focus of attention. This novel form of shared attention lets the cognitive architecture to acquire and correct its perceptual behavior, making it for the operator easier to search for meaningful items in the environment, and processing a video-stream [31]. The computational workload of the cognitive architecture decreases by selective attention since only parts of the scene need to be interpreted (cf. also WP5).

Task T4.3: *Scheduling and balancing communication for cooperation [48 PM]*
(Period: M0-M48, Partners: [TNO/30 PM,DFKI/12 PM,FDDo/3PM, VVFF/3PM], Contributes to objectives: D)

T4.3 investigates how humans want tasks, (shared) information spaces, and dialogues to be adapted for different positions within the CTL-model and how this depends on the user and context. For estimating the user state, this project has a practical approach. The CTL-model describes the task demands in terms of observables (e.g., environmental events and dialogue acts) and human characteristics (e.g., experience and preferences). First, a CTL-model will be generated and validated for this application domain (i.e., the relations between the observables and characteristics with their projection on the three CTL-dimensions on one side, and performance and mental effort of the rescuers at the other side). Based on the user model and context information, the momentary CTL can be estimated, and the information provision and dialogue style can be tailored to this CTL state T4.3 develops mechanisms in conjunction with WP3 to accommodate the dynamic preferences in dialogue production, and planning. There are several mitigation approaches that are explored and assessed (i.e., timing, the actor that has the initiative, feedback to the human actor). Of utmost importance for the envisioned dynamic multi-actor operations and information processes, is to support the maintenance of adequate levels of Situation Awareness (SA) and trust concerning the dynamic task allocation in complex environments [204, 164] and human-robot cooperation [58, 28]. The scheduling and allocation of tasks to the robot and user should be clear for both. A pitfall of delegating tasks is that the SA might decrease and can hardly be regained when needed, resulting in bad performance. By requiring user involvement during the setting-up of working agreements(e.g., defining the conditions under which specific dialogue-modes are active; T4.1), by providing adequate SA-displays and mode-feedback, and by using running commentaries (§1.3.5.3), we keep SA relatively high. An important aim is to prevent the occurrence of the eight demons of SA that Endsley [51] distinguishes: attentional tunnelling, requisite memory trap, stress, data overload, misplaced salience, complexity creep, errant mental models and out-of-the-loop syndrome. The adaptation is further refined by attuning the information sharing and dialogue to the contextual dynamics and possible suboptimal human states (e.g., fatigue). The attuning can be (partly) under control by the human or robot according to the previously determined working agreements (T4.1). Whereas T4.3 strongly focuses on the perspective of the human operator (cooperation demands), a robot's perspective is adopted in T5.3 and T5.4. These WP5 tasks trigger operational demands on shifts in mixed-initiative and attention, to be balanced against the cooperation demands for scheduling and communication modeled here in T4.3.

Description of work – Milestones

Milestone MS4.1: *Task load and selectional attention determination by knowledge and context factors* (M10)
Task load of the user and focus of attention are determined with knowledge and context factors such as, user characteristics, and immediate dangers. The load, attention and user models have face validity (i.e., field operators recognize and confirm the indentified load levels and attention foci). These models provide the foundation for cooperative human-robot load balancing and efficient allocation of attentional resources (i.e., the models distinguish "classes of performance deficits" for setting the conditions of robot's adaptive behaviours).

Milestone MS4.2: *Adaptation strategies to decrease task load and focus attention* (M22)
Several mitigation methods for attuning the human-robot cooperation are identified and tested. The adaptation strategies decrease the risks of cognitive taskload- and attention-based SA failures (such as need for closure and cognitive lock-up). By adequate scheduling of tasks, setting the focus of attention to the appropriate information and objects, and providing the required dialogue styles and timing for the "running commentary", human performance remains adequate under critical conditions.

Milestone MS4.3: *Cooperation strategies to decrease cognitive taskload* (M34)
Using the previous results and the corresponding domain analyses, we build an advice function for establishing adequate adaptation mechanisms. Furthermore, feedback is provided on the cooperation and situation in which the human and robot operate (i.e., situation overviews, performance, cognitive taskload distribution and attention allocation over time). Based on this feedback, the field operator has better SA and improves cooperation with the robot.

Milestone MS4.4: *Working agreements for cooperation strategies* (M44)
For planning and preparation, the field operator can set working agreements for dynamic task allocation (level of robot autonomy) and adaptive dialogues. Adaptation is now an integrated part of planning, preparation, execution and evaluation (debriefing) of USAR missions. In this way, the operator is in-the-loop, is under control and maintains adequate SA, and cognitive load balancing is tailored to the specific user and context demands. Furthermore, feedback is provided in such a way that the operator can learn from the experiences and improve the adaptation settings during the planning of future missions, and the user interface provides an easy-to-use function for cooperative human-robot exploration (generating, adjusting and maintaining a joint

map of the scene and events).

Description of work – Contributions to project

The aim of WP4 is to improve joint exploration by decreasing both the cognitive taskload for the human workers and the robots, contributing to objective 2 (Situated cognitive user models). Models for cognitive taskload and attention require information from the environmental context (WPs 1, 2), interaction context (WP3), and the operational context (WPs 5–6). If there is for instance fire detected nearby the robot or the human, the selectional attention should focus on this threat, and the cognitive taskload of the human probably increases. Directive information from the interaction is in human-human dialogue of major influence on selectional attention; this is the same for our robots. Furthermore, information that is known beforehand, such as what did the building look like beforehand and how many people were approximately in the building, help in directing the attention. Moreover, knowledge about the current interaction, domain and user (because we work with professional users, training is an option) supports in determining the cognitive taskload of this specific user in this specific situation. This WP also provides input for the same WPs it receives input from. The WP as a whole contributes to WP7 to develop a robot for joint exploration. The models of task load and attention provide input for adaptive HRI (WP3) and adaptive planning and cognitive execution (WP5). The focus of attention also helps constructing and disambiguating the environment models developed in WPs 1–2, and focus skill acquisition (WP 5). Finally, dynamic saliency maps provide input for structuring the topological and conceptual map layers (WP1). This WP will provide core UI design and evaluation activities aiming at theoretical and empirical founded solutions to support human's situation awareness and performance. The models of other WPs contribute to the content and means for interaction, which will be evaluated in this WP. Test outcomes, as explained in the next ”evaluation” section, will be fed back into these WPs.

Partner	Contributions
TNO	models of cognitive taskload, user models (domain-specific), knowledge on how context influences user models, dynamic planning and task allocation, cognitive engineering
DFKI	selectional attention mechanisms, domain inferencing and user models, salience ensembles
ROMA	dynamic saliency maps, 3D-projection of perceptual attention, multi-modal perceptual attention
Fraunhofer	real-time aspects of attention (>30Hz), multi-modal perceptual attention, interaction between perceptual attention and object detection & classification (class-based detection)
CTU	interaction between perceptual attention and object detection & classification (class-based detection)

Description of work – Risk management

Probability: Low (unlikely to occur); medium (not unlikely to occur); high (likely to occur)

Gravity: Low (uses available mechanisms within system); medium (uses available technology, not yet in system); high (requires new technology).

Approved by EC on 2009-12-09

Risk	Probability	Gravity	Contingency Plan
Not all three dimensions of the CTL model can be derived from user and context	medium level	low level	The selectional attention models, functional SA, and dialogue help in disambiguing the context and retrieve the relevant information
The SA decreases, because of task delegation	medium level	high level	Establish working agreements and use running commentaries when tasks are delegated
The taskload increases, because of running commentary	high level	high level	Find, by doing experiments, a balance between SA and task load
The selectional attention mechanism is not that selective	high level	low level	Small support to focus or narrow attention is already beneficial

Description of work – Evaluation

To evaluate progress of the WP activities, we need to measure the effects of the cognitive taskload balancing, to assess the adequacy of the attention focus, and to record the improvements on shared situation awareness.

Validation of task load model and user model patterns:
This part of the research concerns the validity of the model predictions on operator performance, knowledge and judgement. For *performance*, we measure the effectiveness (accuracy and completeness) and efficiency (time, cognitive load balancing) of the operations in "prototypical" and critical scenarios according to ISO 13407 (Human-Centered Design Processes for Interactive Systems). The SOWAT tool is used to record Cognitive Task Load [76]. Users get into a "realistic" emotional state for the scenario via appropriate methods [125, 156].

For *knowledge*, we measure Situation Awareness (SA) and its acquisition (i.e., how performance and SA develops over time). SA is an important constraint for excellent performance, capabilities for adapting to new situations (resilience), and learning from the experiences. The evaluation includes all three levels of situation awareness–perception, comprehension and projection– based on the standard methods of Endsley [49, 50, 51].

For *judgement*, we measure an operator's satisfaction, trust and emotional responses via standard questionnaires [237][151]. In general, the evaluation tools provide knowledge about how the interaction should be adapted to balance the cogntive taskload.

Validation of selectional attention model:
The performance of the selectional attention mechanisms can be measured by applying them to annotated USAR scenes. By comparing this performance to the performance of human subjects the biological validity can be determined. In the first phase these are still images, gradually changing to real world moving images.

Evaluation of shared situation awareness:
Shared human-robot SA is important for effective and efficient cooperation. This workpackage assesses the ease-of-sharing information and the effectiveness of complementing, disambiguing and critiquing information of an actor. Measurements include information on the history, current situations, and predictions.

Deliverables:

The software prototypes developed in WP4 serve directly as input to WP7 (and are part of the yearly P deliverables there).

DR 4.2.1 *Validated task load, attention and user model patterns that specify the relationship between task demands and user properties.* R. (M16) [29.5 PM]

DR 4.3.2 *Theory and evaluation of working agreement method and HRI-adaptation to different contexts based on adaptation guidelines, adaptive IU and interaction model.* R. (M34) [37 PM]

DR 4.4.3 *Validated intelligent working agreement mechanism to set-up adaptive HRI.* R. (M40) [15 PM]

DR 4.4.4 *Summative evaluation and theory of the setting-up and usage of adaptive HRI.* R. (M48)[20 PM]

Approved by EC on 2009-12-09

1.3.5.5 WP5: Flexible planning, learning and execution for joint exploration .

Work package number:	3	Starting date or starting event						1
Work package title	Flexible planning, learning and execution for joint exploration							
Activity type	RTD							
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	10	4	9	2	6	5	42	0	7

Objectives

WP5 contributes to project objectives 1,2,4. WP5 provides a cognitive execution model. The model combines flexible planning and cognitive models for execution monitoring with methods for early acquisition and adaptation of skills. Together, the model captures the cognitive level at which operational demands are monitored (in combination with adaptive robot control in WP6). These demands are balanced with the cooperation demands coming from cognitive user modelling (WP4) to yield natural human-robot cooperation (WP3). WP5 formulates three tasks:

A Learning Skills. WP5 investigates skills involved in selecting and coordinating multiple tasks when operating. Skills are learned by continuous interaction with humans, via demonstration and through an action-reaction paradigm, to acquire the effects of actions and processes.

B Cognitive execution and monitoring. Learned skills allow the architecture to arbitrate resources under time constraints and compatibilities among execution and communication processes (operation, cooperation). Cognitive execution constructs a mapping from internal states, shared among several activities, to choices of actions and processes basing decisions on task driven-attention.

C Adaptation. Flexible behaviour planning, morphological adaptation, attentive exploration and switching execution between tasks provide a high performance adaptation to human requirements in USAR mission operations.

Description of work – Tasks

Task T5.1: *Planning activities specification with end user [9PM]*
(Period: M0-M12, Partners: [Roma/3 PM, DFKI 1/PM, Fraunhofer 1/PM, ETHZ 1/PM, VVFF/3PM], Contributes to objectives: A,B)

Specification of learning scenarios, definitions of standards for skills learning (T5.2). Definition of primitive processes for affordance learning. Definition of a dictionary of structures for action-reaction learning. Specification of a task graph to describe priorities on task switching. Hierarchy of activities for setting temporal constraints and compatibilities between processes.

Task T5.2: *Learning skills for functioning processes and task execution [13PM]*
(Period: M0-M24, Partners: [Roma 8/PM, Fraunhofer 1/PM, ETHZ 2/PM, VVFF/2PM], Contributes to objectives: A)

While jointly exploring an area, the NIFTi human-robot team continuously acts and interacts. This makes it possible for them to quickly adapt execution to address sudden needs, which typically arise in a dynamic, unknown environment. T5.2 addresses the problem of how a robot can control task execution in such a dynamic setting. T5.2 develops methods for acquiring the skills necessary for such control. Novel is the use of human attention (as a form of demonstration) in skill acquisition, to learn were to focus attention while a human is performing meaningful tasks.

Learning for cognitive execution is performed offline. T5.2 gathers data on human gaze fixations, by observing one or more people performing specific tasks. From the sequence of fixations meaningful scan-paths are built. These scan-paths show action evolution as a gradual process of change and development of more complex states, in terms of precondition and effects. Scan-path construction is based on clustering fixations [9, 15, 1], defined in a high-dimensional space of features. Features are projected via factor analysis and Gabor based motion classification, into a subspace that produces a saliency map of affordances (in terms of actions scan paths). Offline training sessions are made possible at the end-user sites, and are further replicated in laboratory settings. This guarantees sufficient data on gaze fixations, to provide correct scan-paths.

The saliency map models actions and the classification of their effects, in highly demanding contexts. Observations of successful processes achieved in classified contexts are collected to assess activities networks or work-flow, with time constraints and compatibilities (from mapping to manipulation to terrain adaptation). These causal and temporal relations, with their constraints, are learned, using an observation-state Bayesian framework. Choices are estimated according to specific contexts suitably formulated via priors (see T5.1). This leads to learning interfaces among active processes of the cognitive architecture such as how to switch from one sensor to another (e.g. using the UAV as roving sensor, T6.6), from a current task to one in highest demand. Dynamic saliency maps also provide a basis for modelling attention in cognitive user models (T4.2).

Task T5.3: *Task-driven attention for coordination and communication[13PM]*
(Period: M0-M24, Partners: [TNO 4/PM,DFKI 3/PM, Fraunhofer 1/PM Roma 4/PM, CTU 1/PM], Contributes to objectives: A,C)

T5.3 focuses on learning skills for coordinating human-robot interaction with drives for mixed-initiative and attention, arising from task execution. A typical example is a request from the operator to change path to reach a certain objective using specific sensors, for example in the presence of smoke, and to balance that request with operational demands on adapting morphology. This requires two issues to be addressed.

The first point is to learn how to understand operator indications, connecting them to plans, actions, and their executions. (This uses the specifications for what kinds of instructions an operator can provide, set up with the end user in T5.1). WP5 models this as mapping logical task descriptions (connecting representations from WP3 to those used in WP5, cf. e.g. [23, 130]) to mechanisms for subscribing a service that is used in task execution (T5.4).

Secondly, the architecture needs to maintain task-driven attention towards the current goal and towards human requests. It needs to suitably shift between the two activities (like someone who is driving with a human-navigator at her side). The attention of the cognitive architecture is, thus, focused not only on its current task but also on information the operator has solicited. The architecture needs to evaluate how to pass from one context to the other, how to adapt its strategy. Learning here is online, based on action-reaction and on classification of information communicated by the operator. T5.3 provides the hierarchy of switching criteria in the context of cooperative execution, in particular for mixed-initiative planning. The hierarchy is maintained using the running commentary (WPs 3–4).

The method developed in T5.3 elaborates on non-parametric Bayesian models for learning switching linear dynamic systems [63] considering groups of data generated by different, related processes. Compatibilities and time constraints (cf. T5.4) are considered essential relations.

Task T5.4: *Adaptive behaviours in flexible temporal planning[18PM]*
(Period: M12-M48, Partners: [DFKI 1/PM,Fraunhofer 2/PM, CTU 2/PM, ETH 2/PM Roma 11/PM], Contributes to objectives: B)

Execution involves a continuous coordination loop of all the robot components. It integrates the functioning of different stimuli, movement, morphology adaptation and correct acquisition of information from human-interaction. Execution maps internal states to task performance, to monitor what is needed for performing a specific task with respect to resources, operators, sensors, etc. Adaptation is based on the learned ability to use the correct parameters, the correct balancing of all the components and resources, and exploits planning primitives to afford execution. T5.4 builds on T5.2 and T5.3, focusing on further forms of skill learning.

T5.4 develops methods for adaptive flexible plan execution, using a dynamic temporal model for joint exploration and joint management of the resources at hand (sensors, time allocated for tasks, actuators, internal states, priority list of requests). The methods are based on an integration of logic and probability, based on a BLOG-style [163] representation of processes. For modelling temporal compatibilities and constraints T5.4 exploits Flexible Temporal Golog (FTG)[57] and the Situation Calculus [209, 201]. T5.4 maps learned parameters and features into choices of concurrent processes (flexible time), to adapt behaviours to multi-modal actions and concurrent tasks. Failures in both execution and communication are managed through specific game theoretic strategies (considering the team in its entirety, not just the cognitive architecture as a single actor) ensuring that prior tasks do not stay unexecuted. FTG is both integrated with Matlab and completely executable in C++. The underlying logic is essential for a systematic model but online executability is ensured in the action language.

Parameters, features, constraints and compatibilities, all specify processes that can be built on parameterised actions. The preconditions define when processes can be executed, and the anticipated effects of doing so can be learnt by compiling instructions provided through interaction.

Task T5.5: *Situated exploration history[13PM]*
(Period: M0-M48, Partners: [DFKI 5/PM, Roma 6/PM, Fraunhofer 2/PM], Contributes to objectives: C)

Task 5.5 develops the situated exploration history. The purpose of this history is to provide temporal organisations over grounded planning-, user model and communication content. The history reflects two temporal organisations. One, it reflects a "log"-like organisation, i.e. the order in which content for action planning, action execution, and interactions was constructed and communicated, and what cognitive task load and attention were associated with that content. The history represents content in the logic-based formalism of T5.3, and referentially relate content-representations. Second, the history reflects a temporal organisation over the events in the dynamic, spatio-temporal situation awareness in which that content can be grounded. These organisations are modelled using event structures which represent both temporal and causal aspects, after [165, 17, 44, 130]. Hierarchical relations are added to event structures, to capture different levels of event-based organisation (e.g. the achievement of a goal, and the preparations, actions, and resulting state leading up to that goal). To derive these relations, domain inferences over action/event ontologies are combined with temporal reasoning [128]. Adaptive grounding and communication have access to the history to resolve and ground spatio-temporal references (T3.1, T3.3).

Task T5.6: *Attention-driven exploration[17PM]*
(Period: M18-M48, Partners: [Roma 10/PM,Fraunhofer 2/PM, Blue 2/PM, CTU 2/PM, ETHZ 1/PM,VVFF/2PM], Contributes to objectives: C)

T5.6 puts theory into practice. It implements the effective execution of the planned activities, following T5.1–T5.5. T5.6 takes care of robot attentive exploration, guiding task execution to collect data, to reach positions (e.g. to identify injured people, to gather samples or to verify the presence of some specific object, to gather information requested by an operator). T5.6 ensures the architecture performs the required analysis tasks systematically. The focus on information gathering in a cooperative context distinguishes exploration from navigation or motion planning (these are provided by WP6, T6.3 and T6.4). Attention-driven exploration exploits all the information obtained from the different maps (including the environment models from WPs 1–2 and the saliency maps), from communication with operators (WP 3), from the executive scheduling (jointly with T4.1 and T3.2), and finally by context assessment (morphological adaptation, current visibility etc.).

All this information is suitably used to focus attention on the correct balance between what can be performed and what is needed to be performed. Activities planning is based on urgency and priorities (e.g. [1, 31]). Following the inference of these possible tasks, local goals are established and chosen for execution according to requests. A work-diagram of different concurrent activities is provided to ensure plan integration of the team activities (cf. also T4.3–T4.4). Failures are registered within the team activities and recovery is evaluated along the specified schedule (notably, time constraints and compatibilities among processes).

Description of work – Milestones

The milestones define measurable capabilities to be delivered by this WP. (See below for evaluation metrics.) These capabilities incorporate the functionality provided by the WP tasks, and contribute directly to the milestones for the integrated systems of the roadmap (§1.1.5, §1.3.7).

Milestone MS5.1: *Specification of contexts for learning* (M10)
The basic paradigms for learning the hierarchy of skills necessary to manage the execution, coordination and adaptation are specified and the contexts are defined. These can be revised in subsequent evolution of the system under the need of further constraints.

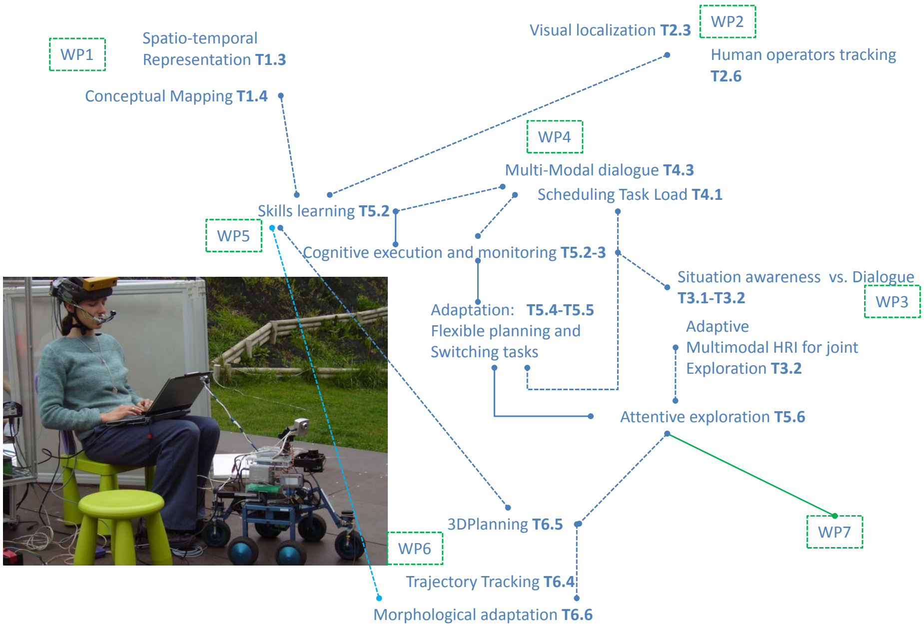
Milestone MS5.2: *Skills are learned for mapping internal states to execution* (M22)
This step overview the basic skills that are needed to coordinate the internal resources of each agent. In this phase the basic cognitive architecture is supposedly designed thus the internal communication between different states , from failures control to perception-execution-communication is achieved by evaluation of the functioning needs. At each step of the operation loop the robot knows what has to be deployed, in terms of resources to activate as requested.

Milestone MS5.3: *Skills are applied to cognitive execution functioning and planning* (M34)
Each agent is able to plan and execute different tasks, according to specific requests, and thus to measure both time constraints and compatibilities between resources and processes. In this phase each robot is able to establish to what extent communication and cooperation can be satisfied by switching between two or more tasks in order to comply with specified urgencies.

Milestone MS5.4: *Adaptive behaviours are used for planning short-time operations jointly with rescuers* (M44)
At this step switching is developed to a further degree of autonomy. The robot can interact with in-field rescuers to verify mixed initiative. During the execution of a task the robot is requested to switch and to remap internal states to current context (by in-field rescuers). The robot is able to deploy an up to date execution strategy that subsumes the new context and the operators requests or to deploy a fresh strategy of step by step interaction.

Description of work – Contributions to project & state-of-the-art

WP5 develops a novel concept of execution. Cognitive execution is meant to comply with real adaptation to changing objectives, following operators instructions (e.g. dialogue, commands) that require switching between tasks and flexibly revise plans and processes. This is crucially based on the ability to learn several skills using attention and generated gaze scan-paths that show the affordances of processes, including the communication steps, at several levels of details (from activities to HRI). Learning skills provides classes of parameters and features for choosing strategies of actions according to time and compatibilities constraints. So far no attempt has been made to create new primitives and parameters on line (on line adaptation) based on the robot's experience and interaction. This requires a novel formal combination of declarative programming (such as Golog scripts), based on logic descriptions, with statistical relational learning, and probabilistic reasoning.



The figure above shows how the main functionalities developed in WP5 are connected to the other WPs tasks. The strong interaction of WP5 with most of the tasks is due to the need to learn the skills for execution monitoring and to correctly balance activities to deploy cognitive control for joint and adaptive operations.

Partner	Contributions to WP5 efforts
DFKI	situated spoken dialogue processing for HRI, multi-modal dialogue, intelligent user interface, adaptive mechanisms for reference resolution and grounding, situated exploration history
TNO	interface design for USAR, intelligent user interface, multi-modal dialogue, shared control, situated exploration history
Roma	Attention based learning, Saliency Map construction, flexible planning and execution monitoring, adaptive mechanisms for planning and execution monitoring, shared control
Fraunhofer	flexible planning and execution monitoring, shared control, situated exploration history

Description of work – Risk management for WP5

Risk	Probability	Gravity	Contingency Plan
Difficulty in skills discrimination	high level	medium level	Hierarchy of skills and association graph between states mapping and skills, between primitive tasks and skills, given the hierarchy and the associated graph build the appropriate discrimination.
Difficulty in mapping internal states to an execution network	medium-high-level	medium level	Test each state, and estimate resources and activities by local Bayes.
Difficulties in plan adaptation	high-level	high level	Short operations, strong mixed initiative, find the set of models satisfying the compatibilities and then retry fitting.

Description of work – Evaluation metrics

There are three levels of evaluation of the WP activities. The first level consists in verifying the ability to identify, via the learned skills, the functioning and resources needed at each step of a single operating loop (Task 5.2). The second level concerns the reliability of execution, from recovering from failure to interaction with remote and/or in-field operators (Task 5.4). The final level is the quality of execution that involves task switching, time constraints, mixed-initiative and ability to subsume urgencies in contingent plans by flexibly adapting to changing contexts (Task 5.6). All the tests will be performed in a NIST-like arena level orange and red (final tests) within a single cycle operation (a specified task in the arena).

Efficiency

Efficiency measures the level of acceptability of the executive control. Where the level of acceptability depends on the ability to guarantee execution in the correct timing and resources availability.

Measures for efficiency and acceptability (in a single cycle operation):

1. misalignments between internal representation and sensor readings.
2. Plan latency.
3. Execution latency.
4. Mean values of execution soft recoveries.
5. Mean values of execution hard recoveries

Effectiveness of adaptability

Effectiveness is a measure for task execution and it is recorded within an operation cycle. Effectiveness of task execution is evaluated with respect to the internal resources used and the number of interventions needed (shared control) and failures recovered.

Measures for effectiveness and adaptability (in a single cycle operation) :

1. number of critical failures.
2. Failures requiring operator intervention.
3. Number of failures recovered.
4. Number of in-time switches between components processes.
5. Acceptance of in-line commands (robot-operator interaction).
6. Time adequacy and plan optimisation (mean time idle).

Efficiency of flexible planning in attention driven task

Task efficiency: is measured in terms of the number of interesting regions visited, of plan steps required to achieve a goal (including re-planning of actions), with respect to the percentage of goal achieved. In synthesis attentive task execution is efficient if resources are optimised with respect to goal achievement.

Measures for efficiency in attention driven task (in a single cycle operation) :

1. mean distances 3D saliency map and objectives.
2. Number of interesting regions visited.

3. Time spent in interesting regions.
4. Roc curve of interesting points reached.
5. Error estimation for effective fixations.
6. Errors in switches to operator commands.

Robust fitting of learning abilities
Evaluation of robust fitting concerns measuring how the correct classification for the three main aspects of flexibility, have been learned. If the performance measures for efficiency, effectiveness of adaptability, satisfiability of task execution and efficiency of flexible planning are fulfilled then robust fitting is achieved. Thus time constraints, compatibility of processes and communication, and mapping internal states to execution, is valid and comprehensive of any context.

Deliverables:

Domain analyses describe field experiments with the gaze machine for gathering fixations and gaze trajectories. The software prototypes developed in WP5 serve directly as input to WP7 (and are part of the yearly P deliverables there).

DR 5.1.1 *Domain analysis and specifications: context scenario and skills primitives.* R. (M10) [9PM]

DR 5.1.2 *Methods and paradigms for skill learning based on affordances and action-reaction observation.* R. (M12) [11PM]

DR 5.2.3 *Hierarchical structure of learned skills, scan paths, saliency map of activities and communication interfaces.* R. (M24) [15PM]

DR 5.3.4 *Resources management and mapping from states to execution, integrating linear dynamic models learning into theory of actions..* R. (M30) [18PM]

DR 5.3.5 *Flexible planning with time constraints and compatibilities.* R. (M36) [13PM]

DR 5.4.6 *Mixed initiative planning and user requests subsumption. Adaptable strategies for complying to robot-team and users requests.Adaptable strategies for complying to robot-team and users requests..* R. (M48) [19PM]

1.3.5.6 WP6: Adaptive operation .

Work package number:	6	Starting date or starting event						1
Work package title	Adaptive operation							
Activity type	RTD							
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	0	0	20	18	27	0	8	0.25	0.25

Objectives

WP6 provides the functionality for low-level adaptive control, including adapting locomotion and perception morphology. WP6 has the following objectives:

(A) Develop a novel robot platform that outperforms existing vehicles currently used for USAR missions. The main effort is made towards the extension of the climbing capability and the weight optimization.

(B) Develop adaptive control for the designed platform, including the use of a UAV as roving sensor.

WP6 contributes to achieving Objective 4. For the UAV low-level control, a commercial system is used.

Description of work – Tasks

The WP6 tasks address the objectives. T6.1 identifies the needs of the end-users and the partner requirements. Based on these specifications, T6.2 takes care of the design of the platform for an Unmanned Ground Vehicle (UGV) and Unmanned Aerial Vehicle (UAV), prototyping and sensor integration. T6.3 tackles the trajectory execution and the traversability analysis, whereas T6.4 and T6.5 focus on high level planning and semi-autonomous navigation. T6.6 addresses adapting perception morphology, using the UAV as roving sensor.

Task T6.1: *Platform specification [4.5 PM]*
(Period: M0-M2, Partners: [BLUE (UGV)/1 PM, Fraunhofer (UAV)/1 PM, ETHZ/1 PM, ROMA/1 PM, FDDo/0.25PM, VVFF/0.25PM], Contributes to objectives: A)

T6.1. establishes a clear specification of the platform in terms of size, weight, climbing capabilities and sensing. The requirements come from the end-users (rescue organizations) and from the other partners of the project. .The exact specification of how the platform capabilities are to be realized, needs to strike a balance between the user- and partner requirements, and the scientific goals of combining active solutions for variable morphology (typical for military applications) together with lightweight passive structures with high climbing abilities (typical from space applications), and adding sensors permitting as much autonomy as possible given shared control with a user. The possibility to integrate an arm on the platform is investigated. Such a device is essential for deploying probes and moving obstacles blocking the path. Similarly, a 3D range sensor is required for acquiring spatial models of the environment that are used by almost all tasks of the project (localization, situation awareness, autonomous navigation, GUI's etc.). Furthermore, the requirements for the UAV are specified.

Task T6.2: *Platform design and sensor integration [23 PM]*
(Period: M2-M12, Partners: [BLUE (UGV)/12 PM, Fraunhofer (UAV)/4 PM, ETHZ/5 PM, ROMA/2 PM], Contributes to objectives: A)

The design of the platform starts as soon as the specification is ready. As mentioned in §1.2.8, the weight, power consumption, control complexity and climbing performance are optimized using hybrid locomotion. The novel platform combines the characteristics of active and passive locomotion concepts. The developed concept is presented to the partners and end-users for validation before it is actually manufactured. The most sensitive components of the platform will be identified and the actions taken to limit the risk of failure presented in the platform's documentation. A functional platform with integrated sensors, software interface and embedded computers is made available by the end of the first year. A total of 7 platforms will be built. After Yr1 delivery, the platforms will be iteratively updated to in keeping with the development cycle (notably, user requirements and user experiments). The UAV is based on commercially distributed kits and components, to be available at M6. The choice of the exact UAV platform is subject to requirements (T6.1). An increasing number of companies provides suitable platforms, e.g. microdrones from Ascension Technologies, microdrones, or HiSystems.

Task T6.3: *Trajectory control and traversability analysis [12 PM]*
(Period: M12-M24, Partners: [BLUE (UGV)/2 PM, ETHZ (UGV)/6 PM, Fraunhofer (UAV)/2 PM, ROMA/2 PM], Contributes to objectives: A)

T6.3 is dedicated to trajectory control. The great majority of path trajectory control laws for either kinematical or dynamical mobile robot models are designed assuming ideal actuators, i.e. assuming that any commanded velocity or torque (in the kinematical and dynamical cases respectively) will be instantly implemented regardless of its value [42, 95, 96]. Real actuators are far from being ideal and obstacles blocking the trajectory should be considered. In particular, only bounded velocities and torques can be realized for any given command. With reference to the kinematical model of the new rover robot, a known path following control law is modified to account for actuator velocity saturation [136, 97]. The implemented solution is particularly useful for rough terrain applications where accounting for actuator velocity saturation may have a large influence on performance.

The 3D models and the localization developed in WP1 are used to support this task. The terrain classification for rough-terrains presented in [20] is implemented and adapted to the NIFTi platform. T6.3 focuses on the analysis of projected trajectories based on the kinematics constraints of the rover, 3D spatial information and soil types. The types of soils are assessed using the approaches described in [26, 36]. Finally, each trajectory is assigned a traversability value that is used to a) select the best trajectory when operating in autonomous mode b) trigger human intervention in case no trajectory is safe enough (WP3). The second point is used to support and test shared control (WP3–5). Furthermore, the interfaces for data transfer between UAV and UGV, 3D models and GUI components (WP3) are developed and implemented. From Task 6.3 onwards, ETHZ leads efforts in developing control of these rovers in defeating terrain configurations, using BLUE's locomotion concept and sensor-interfaces provided by Fraunhofer.

Task T6.4: *3D planning and semi-autonomous navigation [10 PM]*
(Period: M24-M36, Partners: [BLUE/1 PM, ETHZ/6 PM, Fraunhofer/3 PM], Contributes to objectives: B)

Semi-autonomous navigation in a static environment cluttered with 3D obstacles is implemented for the UGV and UAV. At this stage, it is expected that the environment is mostly flat with island of localised obstacles. [228] and [91] are used as the basis for the implementation of the UGV. The 3D planner makes use of the traversability map established in T6.3 to plan the best route towards a goal given by the operator. This planner is used by the high level planner developed in WP5 to a) execute local motion of the platform b) check if a location is reachable and evaluate the path cost.

Task T6.5: *Trajectory tracking in Dynamic Environment [14 PM]*
(Period: M36-M48, Partners: [BLUE/1 PM, ETHZ/9 PM, Fraunhofer/4 PM], Contributes to objectives: B)

This task extends the functionality of T6.4. It enables automatic replanning in case dynamic changes occur in the scene and tackles a full 3D environment. This task relies on being able to detect changes, which is an output of WP1. Approaches such as Field $D*$[56] or $E*$[198] are the basis of the tools developed in this task. These approaches allow to update the output of a path planning stage with new perceptual data, without regenerating the full path. In order to account for kinematic or dynamic constraints, more sophisticated approaches such as nonholonomic trajectory deformation [88] can also be applied.

Task T6.6: *Morphology-adaptation and active perception [10 PM]*
(Period: M12-M48, Partners: [Fraunhofer/6 PM, BLUE/1 PM, ROMA/3 PM], Contributes to objectives: B)

The UAV is used as a roving sensor around the platform. The platform processes the visual input stream from the roving sensor. Decisions for actively deploying the UAV as roving sensor is given as a traversal trajectory for navigating in the environment. If perceptual classification of spatial structure (WP1) and/or potential threats (WP2) encountered on the trajectory indicate that navigation may be impeded, but perceptual information available to the rover is too uncertain or too incomplete to confirm, the UAV is deployed (cf. WP5: T5.3, T5.6). The platform actively guides the UAV using 3D spatial data (from 3D scanner), initial position and pose (camera pose) for the UAV. Once deployed, further processing of the UAV's visual input stream (WP2) combined with the map (WP1) and dynamic attention (WP4) serve to direct further active perception with the UAV. As visual output is communicated to the user, the user also has the possibility to intervene / further direct the deployment of the UAV.

Description of work – Milestones

Milestone MS6.1: *Platform prototyping* (M10)
During the first year of the project, the focus is mainly on hardware development and integration. A functional platform with integrated sensors, software interface and embedded computers is available by the end of the first year (UGV and UAV). The platform can be remotely controlled by an operator and successfully reach a goal in an environment comparable to a NIST USAR Arena Yellow. The UAV with a camera is available and navigates semi-autonomously based on inertial sensors around the robot or on GPS data outside. This milestone contributes to objective (A).

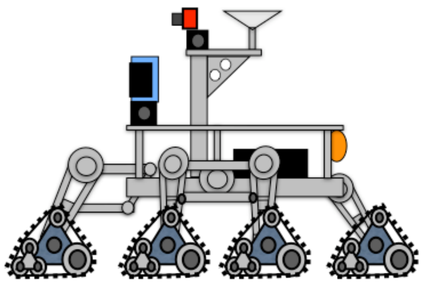
Milestone MS6.2: *Trajectory control and traversability analysis* (M22)
The second year is dedicated to trajectory control for the UAV and UGV. This work focuses on trajectory execution and on the analysis of projected trajectories based on the kinematic constraints of the rover and 3D spatial information. The approach is validated using a predefined set of NIST USAR Orange-level obstacles (terrain patches) placed on a flat ground. For each obstacle, different trajectories are projected, analyzed with respect to traversability, and executed. It is expected that the rover successfully executes the trajectories that are classified as feasible and fails to execute the ones that have a low traversability value. Unclear or negative traversability situations lead to the deployment of the UAV to support additional visual streams around the rover. The local UAV trajectories are calculated based on the rover's 3D laser scanner and submitted to the UAV. The UAV then executes the trajectory. This milestone contributes to objective (B).

Milestone MS6.3: *Semi-autonomous navigation* (M34)
Semi-autonomous navigation in a flat environment cluttered with the same obstacles as used for the previous milestone is implemented during year 3 for the UGV. The platform is able to propose a feasible path to the user using situational awareness data or request attention if a blocked situation is detected along the path. After deploying the UAV and processing the visual data streams the UAV is able to return to the UGV and lands there based on tracked visual markers. This phase requires 3D planning as the planned path may force the robot to climb over obstacles, and guidance of the UAV. This milestone contributes to objective (B).

Milestone MS6.4: *Semi-autonomous navigation with dynamic changes* (M44)
An extension of semi-autonomous navigation to a full 3D environment, up to NIST USAR Red level, is proposed for year 4 for the UGV. Furthermore, modification of the environment occurs after the motion planning and especially during motion execution. The platform has to plan paths in a 3D environment and, based on detected changes (WP1), react to the dynamic event by replanning or requesting human intervention or deploying the UAV. The interaction of the UGV and UAV is improved and optimized. This milestone contributes to objective (B).

Description of work –Contributions to the project & state-of-the-art

The platforms developed in WP6 provide the basis for the sensor platforms for WPs 1 and 2, and the integrated systems constructed and evaluated in WP7. The planning facilities for trajectory control and execution interact with the navigation planning in WP5, and multi-modal communication (WP3). The mechanisms for shared control are integrated with the higher level mechanisms of WPs 3–5. The resulting UGV platform combines passive and active locomotion in an adaptive, variable morphology. The picture below gives an impression of a possible design.



Partner	Contributions
ALL	Specification input, sensor requirement and interface
Fraunhofer	Autonomous UAV control, sensor integration, UAV control interface
BLUE	Develop the novel locomotion concept, integrate the sensors and provides the low level control
ETHZ	Develop the high level locomotion control (obstacle negotiation, autonomous navigation)
ROMA	Best route planning, trajectory monitoring and adaptation to environment changes

Description of work – Evaluation metrics

Locomotion capabilities

The locomotion capabilities are evaluated using the NIST Arena standards. We define three levels of difficulties: easy/yellow, medium/orange and difficult/red. The locomotion capabilities are evaluated using the time needed to explore an area. Comparative evaluations are performed using different terrain difficulties and robots. The goal is to show the curve of degrading locomotion capabilities when non-hybrid architectures are used.

Adaptive control

The level of autonomy of the robot is evaluated using the following human factor measures:

- efficiency, the number of human interventions per time unit and the time needed to execute a given task
- satisfaction, how high users score the use of adaptive control (in conjunction with WPs 3–5)

Description of work – Risk management for WP6

Probability: Low (unlikely to occur); medium (not unlikely to occur); high (likely to occur)

Gravity: Low (uses available mechanisms within system); medium (uses available technology, not yet in system); high (requires new technology).

Risk	**Probability**	**Gravity**	**Contingency Plan**
Inaccurate specification	medium	medium	Make sure that the interaction with users takes place intensively.
Delay in the specification	high	medium	Organize one/two user workshops to collect and validate the specification.
Platform performance inconformity	low	medium	Perform compliance control and test planning at every stage of the development (design, sub-assemblies production, platform assembly).

Deliverables:

DR 6.1.2 contains the API documentation for the platform and software packages. In particular, the API provides access to platform diagnostics:

- *Actuators state (operational, motor saturation, damaged driver, blocked degree of freedom, etc.)*
- *Sensors state (operational, measurement saturated, no data, etc.)*
- *Platform state (operational, safety loop opened, etc.)*

This diagnostics is used in WP5 for task switching.

DR 6.1.1 *Platform specification and design.* R. (M6) [4.5PM]

DR 6.1.2 *Platform manufacturing and sensor integration.* P, R. (M12) [23PM]

DR 6.2.3 *Trajectory analysis: principle and evaluation.* R. (M24) [12PM]

DR 6.3.4 *User interaction and trajectory planning in unstructured environment based on 3D perceptual data: principle and evaluation.* R. (M36) [10PM]

DR 6.4.5 *Trajectory planning in dynamic unstructured environment based on 3D perceptual data: principle and evaluation.* R. (M48) [24PM]

1.3.5.7 WP7: Integration and evaluation .

Work package number:	7	Starting date or starting event						1
Work package title	Integration and evaluation							
Activity type	RTD							
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	20	20	25	4	20	18	14	3	3

Objectives

The objectives of WP7 are to

(A) Do a deep domain analysis of USAR environments with end-users, starting at project beginning and reiterating and refining analyses after every evaluation.

(B) Integrate WP components into a single cognitive architecture, and evaluate on system-level.

(C) Perform end-user evaluations of the integrated systems.

NIFTi explicitly addresses the complexity of integration and evaluation, through integrated methodology, infrastructure, and management (§2.1). Involving end-users in the development cycle, NIFTi ensures effective human-centered integration and -evaluation, and applicability of its results outside lab settings.

The development cycle is also used to feed the WP7 results back to WPs 1–6. It presents the principle means for iterative design and development. Analyses of the end user evaluation (reported in WP7 deliverables) are combined with the next phase of adapting and determining end user requirements in the development cycle. This ensures that the results inform the next phase of development at component- and system-level. The use and possible further contributions to standard robotic benchmarks[5][6] ensure high quality implementation. WP7 yields integrated cognitive robot systems for joint exploration (§1.1.5).

Description of work – Tasks

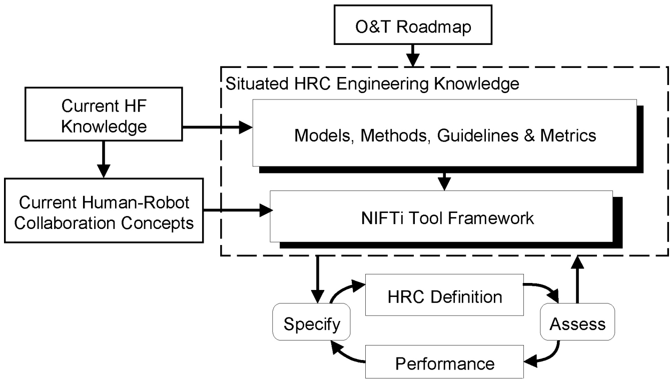
Task T7.1: *Cognitive engineering and domain analysis [22 PM]*
(Period: M0-M48, Partners: [TNO/8 PM, ETHZ/5 PM, ROMA/4 PM, Fraunhofer/5 PM], Contributes to objectives: A, B, C)

For the development and evaluation of human-robot cooperation T7.1 uses the situated cognitive engineering+ framework (CE+) [175]. This framework has been developed and applied in defense and space domains for the design of cognitive support that enhances the capacities of teams and team members during critical and complex operations, both for near- and far-future systems to improve e.g. task load management, trouble-shooting and situation awareness within a user-centered design framework [181, 179]. The figure below provides an overview of the approach.[7]

[5] http://www.robot.uji.es/EURON/en/index.htm

[6] http://wiki.robot-standards.org/index.php/Benchmarks

[7] (O&T = Operation & Technology; SA=Situation Awareness, DM=Decision Making, HRC=Human-Robot Cooperation



The NIFTi cognitive architecture framework plays a central role in the development methodology. The technical development of the infrastructure for the cognitive architecture framework is performed in T7.2. The system generations outlined by the roadmap in §1.1.5 are developed in Tasks T7.3–T7.6. These tasks address the integration and evaluation of the NIFTi framework and its system instantiation, each year.

According to the CE+ model we start with a general concept and implement its components incrementally. A concise roadmap of both the operations and technology, based on the project roadmap (§1.1.5) and the use cases, define the successive steps of the cooperative system to consider. Associated with the consortium (through subcontract) are three end-user organizations from the USAR domain. The end users are involved in establishing use cases and detailing system requirements during example use cases. The example use cases provide information on the the task itself (how is it performed, required situation awareness), end user psychology (e.g. task load, why to attend to what, why certain decisions are made, collaboration), and rescue organization versus individuel rescuer. Field-experiments with the end users iteratively adapt the roadmap where needed to match user requirements. Every year, the end-users are involved:

- 2nd/3rd month to refine and validate the research outcomes (e.g. scenarios, conclusions).
- 6th/7th month to discuss progress and interim results and plans
- 10th/11th month to evaluate and refine the test plan and to conclude in the evaluation

Task T7.2: *Development of NIFTi cognitive architecture framework [35 PM]*
(Period: M0-M48, Partners: [Fraunhofer/12 PM, DFKI/4 PM, TNO/4 PM, Blue/4 PM, ETHZ/5 PM, CTU/4 PM, ROMA/2 PM], Contributes to objectives: A, B, C)

For the development of the "NIFTi framework" of the figure above, NIFTi develops several generations of hardware and software systems. This may include the use of novel state-of-the-art hardware and software systems from outside the project.

To handle the dynamics and the resulting integration complexity, several arrangements and procedures will be defined. The complete software development framework is set up in a virtual machine (VM). The VM is set up and maintained by an experienced partner (Fraunhofer), and is distributed through the project portal. This reduces software installation and update efforts. The VM is weekly updated, and new VMs are set up to incorporate relevant operating system upgrades (for UNIX- and Windows-based platforms). For project-internal software development, several agreements are defined. Eclipse with its several plug-ins is the main software development framework, C++ and Java the main programming languages. An object broker system is used to provide the functionality for combining and distributing different services and functionality. Time critical task are evaluated on real hardware e.g. for the UAV.

NIFTi partners jointly define coding standards, using SVN as revision control system and `doxygen` for code documentation. Developer Wikis on the project portal support bug tracking and further documentation. All software is strictly modular, and hardware-independent modules can be compiled on UNIX- and Windows-based platforms. For each module, the developing partners supply a sufficient unit test for regression testing. Hardware control is capsulated and hidden by an abstraction layer, to support multiple hardware platforms. The coding standard and documentation for each module are dynamically reviewed by involved partners. Following the agile approach short but regular stand-up meetings of the developers via VoIP are initiated to discuss current efforts, plans, and questions. This ensures the timely solution of outstanding problems, and synchronization of efforts. The development methods and standards are discussed and trained at the beginning of the project in a 1-week workshop. Development experience, problems and changes during the project are discussed in a developer workshop every year. The planned student exchange program will further assist integration and knowledge exchange.

Two approaches are used to integrate hardware into the developed systems. For one, WP7 integrates with commercial state-of-the-art platforms, like the VolksBot [94], and Pioneer P3-AT, which the partners already own. Second, integration is done with the NIFTi platforms (WP6).

Task T7.3: *Human-instructed exploration[15 PM]*
(Period: M0-M12, Partners: [all, DFKI/4 PM, TNO/2 PM, Fraunhofer/2 PM, ETHZ/2 PM, CTU/3 PM, ROMA/2 PM, FDDo/0.75PM, VVFF/0.75PM], Contributes to objectives: A, B, C)

T7.3 develops an integrated cognitive robot system for human-instructed exploration. Cooperation involves a human instructing a robot how to explore an environment of NIST USAR Yellow level. The robot autonomously navigates, executing the exploration plan, and communicating what it sees. Central issues for evaluating the human factor in this setting focus on user modeling, assessing variation in task load and attention when building up situation awareness using the robot's experience. The robot needs to completely navigate the area in under 10 minutes. This time criticality imposes requirements for the efficiency of, for example, the robot control and/or robot automation, but also on, for instance, the ability of the user to built up situation awareness and stay cool under the time pressure.

Specify scenarios and use cases. Based on document analyses (there is a rich set of material on USAR missions), interviews and field observations, T7.3 derives a task model that gives an overview of the "core" USAR activities, critical events and environmental conditions. The task model is a common reference for specifying scenarios and use cases (forming the so-called "design rationale"). Task, scenario and use-case analysis is an iterative process. For the Yr1 scenarios, T7.3 works out and tests the following mission in detail; interactively and (semi-)autonomously explore the arena; go-and-return to or go-and-stay at human-identifiable locations in the NIST-arena (see below). During the project, the specifications are refined, maintained and shared throughout the project partners and end-users.

Specify system design. To establish an integrated cognitive system, T7.3 brings together the different types of functionality required for interactive exploration in the semi-structured environments. The interaction between functions is studied using a combination of off-the-shelf technology and new components from NIFTi. The robot is able to acquire semi-autonomous 2.5-dimensional spatial models for situational awareness of partially unknown dynamic environments (WP1). The models capture spatial organization, (visual) landmarks, and terrain classification for traversability and threat analysis (WPs 1–2). The models are layered to enable the connection with qualitative representations to comprehend situated dialogue (WPs 2–3). HRI focuses on instructing a robot where to go to, which crucially relies on 3D-spatial referencing, requiring common ground in understanding landmarks and spatial organization (WP3). Cognitive user modeling focuses on defining models for cognitive task load and attention, to capture the cooperation demands for planning and communication (WP4). Attention and mapping information feed into skill learning for flexible planning, and cognitive execution during exploration (WP5). The system maintains a situated exploration history (WP5). The cognitive architecture is integrated

Approved by EC on 2009-12-09

with a first robot prototype (WP6).

Assess technical capabilities. T7.3 investigates what a robot needs to understand about its environment, to be able to communicate with a human about it in real-life; investigate how a robot needs to understand the environment, to navigate in it; and, investigate what are suitable morphologies for operating in semi-structured environments. The environment complexity is comparable to NIST USAR arena Yellow (multiple levels, different flooring materials, static hazards (e.g. holes), varying lightning conditions and passage way sizes). System elements that are tested in detail are: Spatial mapping (localization, multi-level abstraction, use in navigation), sensor-processing (map-based landmark identification; terrain classification for mapping; odometry; sensor-fusion), and morphology (platform, forms of locomotion).

Assess with end-users. T7.3 assesses how far users are able to instruct the robot where to go. This is assessed in a NIST USAR arena Yellow, by using a tele-operated robot and the robot developed in the project (the robot is required to completely navigate the area within 10 minutes). The experiments test the requirements that users have to trust the robot to rely on its autonomy, keep high SA, and keep low, but not too low, cognitive task load, without degredation in the performance. Furthermore, intermediate results are discussed with the end-users.

Task T7.4: *Human-assisted exploration [15 PM]*
(Period: M12-24, Partners: [all, DFKI/4 PM, TNO/2 PM, Fraunhofer/2 PM, ETHZ/2 PM, CTU/3 PM, ROMA/2 PM,FDDo/0.75PM, VVFF/0.75PM], Contributes to objectives: A, B, C)

T7.4 builds an integrated cognitive robot system for human-assisted exploration. Cooperation extends to mixed-initiative interaction for human-assisted exploration. Environment complexity increases to NIST USAR Orange level with dynamic threats. The robot navigates autonomously, using attention to drive exploration. Central issues concerning the human factor focus on user adaptation, assessing when and how autonomy, communication, and execution should be adapted. It is not a good idea that the robot immediately adapts when the user is feels, for example, a little more tired than usual. The adaptation should take into account the history of that user (e.g. range of level of tiredness) and adapt only when a certain trend is seen for some time (tired for, for example, 30 minutes and on stress there should be a shorter reaction time before adaptation). The definition of stability of the state of the user will be different for every user (e.g. some will have a very fluctuating heartrate, while others are pretty constant) the adaptation of the robot will depend on this and it will be more responsive when someone with a stable heartrate shows a sudden increase in heartrate than when someone with a strong fluctuating heartrate shows a sudden increase. The general hypothesis tested this year is that better adaptation yields better overall performance of a human-robot team.

Specify scenarios and use cases. The 2nd year scenarios exceed the 1st year scenarios in robot functionality, both in HRI as in moving autonomously through an environment. The specification of the scenarios and use cases will be in collaboration with the end-users taking into account the results of the first year.

The 2nd year works out and tests the following mission in detail: interactively explore the arena; autonomously go-and-return to or go-and-stay at human identifiable locations in the arena. The human merely assists in navigation in contrast to the human-instructed navigation of Yr1, resulting in higher demands on robot adaptation and -autonomy.

Specify system design. A robot with increased autonomy acquires semi-autonomously 3D multi-layered spatial models for SA (WPs 1–2). Attention-driven exploration is used (WP5), with the possibility of the user to intervene or instruct the otherwise autonomous exploration. Results from Yr1 about SA, HRI, and variable morphology (WP6) are integrated in the system design specification. Communication focuses in T7.4 on cooperatively and adaptively establishing exploration plans for the robot, including referencing to abstract spatial entities (e.g. unknown locations) and spatiotemporal events (e.g. future actions) for navigation (qualitative characterization of waypoints) (WPs 3,5). Cognitive user modeling focuses on mechanisms for adaptation (WP4).

Assess technical capabilities. T7.4 investigates when a robot needs to use mixed-initiative dialogue, a degree of autonomy sufficient to explore a semi-structured environment with only communicated assistance, and a variable morphology which complicates its sensor integration. The environment complexity is comparable to NIST USAR arena Orange (multi-level, different types of flooring material, present static hazards (e.g. holes) and dynamic ones (collapsing structures), varying lighting conditions, and passage way sizes). System elements that are tested in detail are: Spatial mapping (localization, multi-level abstraction, use in navigation), sensor-processing (visual objects, landmark recognition, terrain classification, and perception of changing environmental aspects), morphology (platform, forms of locomotion).

Assess with end-users. The assessment with end-users have the same focus as in year one, how do human and robot understand each other and does the new interaction influence the trust, performance, SA, and cognitive task load. The robot is tested against the robot from last year and a tele-operated one in an Orange arena, which should be navigated without time-limits. The human tasks during the experiment are more extensive than in Yr1, because the robot has more autonomy. Furthermore, instead of uni-directional interaction from human to robot there is now communication between human and robot, setting further requirements for performance and ease of use.

Task T7.5: *In-field joint exploration planning [17 PM]*
(Period: M24-36, Partners: [all, DFKI/4 PM, TNO/2 PM, Fraunhofer/2 PM, ETHZ/3 PM, CTU/4 PM, ROMA/2 PM, FDDo/0.75PM, VVFF/0.75PM], Contributes to objectives: A, B, C)

T7.5 builds an integrated cognitive robot system for in-field joint planning of exploration. An in-field rescuer communicates with a robot to establish a joint exploration plan. The human factor studies the effects of cooperation on building up situation awareness. The hypothesis is that the running commentary facilitates human-robot collaboration in field, lowering cognitive task load while simultaneously improving the human's situation awareness. Particular focus is on dealing with perceiving and describing the environment from different perspectives, adaptation to cognitive task load, and coordinating ongoing (individual) exploration tasks.

Specify scenarios and use cases. HRI now includes an in-field rescue worker. The scenarios and use cases in the 3rd year will be more focused on perspectivization. How do the robot and rescuer communicate about objects from their own and each others perspective? This furthers the development of the robot, by increasing its abilities to ask for and give information. By developing the use cases we also look at how to provide the user with the best way of working taking into account stress factors .

Specify system design. Because both the robot and the user are in-field during deployment, the SA is extended to handle understanding from the environment from different perspectives, both object recognition as understanding the situation from the human's perspective.

Assess technical capabilities. T7.5 investigates what a robot needs to extend spatiotemporal understanding, action and interaction to deal with the perspectives of an in-field rescuer, when jointly planning exploration. HRI and planning can adaptively understand, and produce, perspectivized descriptions when establishing a joint exploration plan (WPs 3, 5). Adaptive, multi-modal HRI extends the work on jointly establishing exploration plans for the robot in the context of USAR field practices (WPs 3–5). Furthermore, the variable locomotion and perception morphology is extended to deal with environments of NIST USAR Red complexity (WPs 5,6). System elements that are tested in detail are: Spatial mapping (localization, multi-level abstraction, use in navigation; WP1), sensor-processing (multiview of visual objects, landmark recognition, terrain classification, and perception of changing environmental aspects; WP 2), and morphology (platform, forms of locomotion and perception; WP6).

Assess with end-users. The integrated robot is assessed by end-users performing experiments with the new robot, the one from last year, and a tele-operated robot, in a NIST USAR arena with Red complexity (without time-limits). The tasks are focused on informing each other about the situation using knowledge about each others perspective and the appearance of objects from different perspectives. In addition to performance, trust, user state (task load and emotional) are an important focus of evaluation.

Task T7.6: *Sharing situation awareness [17 PM]*
(Period: M36-48, Partners: [all, DFKI/4 PM, TNO/2 PM, Fraunhofer/2 PM, ETHZ/3 PM, CTU/4 PM, ROMA/2 PM, FDDo/0.75PM, VVFF/0.75PM], Contributes to objectives: A, B, C)

Cooperation extends interaction with in-field operators, investigating how an operator can cooperate with a robot to build up a shared situation awareness. The human factor focuses on active cooperation, investigating working agreements to establish how a shared awareness can be efficiently build up during joint exploration. Working agreements further robot behavior and communication adaptivity, to fit in with human practice. Environment complexity remains at NIST USAR Red level. Evaluation uses task-dependent time-limits.

Specify scenarios and use cases. In the 4th and last year the focus lies on extending situation awareness and multi-modal HRI to share situation awareness between robot and rescuer. This holds for the scenarios and use cases in that they highly resemble current practices with human-human rescue teams. Movement of the robot and interaction with the robot should be efficient and effective.

Specify system design. T7.6 extends interaction with in-field operators, investigating how an operator can interact with a robot to build up a shared situation awareness. T7.6 develops novel methods for acquiring and maintaining situation awareness using the robot's own experience, and communicated experience

Assess technical capabilities. Environment complexity remains at NIST USAR Red level. The robot adapts to, and integrates, situational information of differing spatiotemporal referential nature, at different levels of detail, trying to understand how the environment may be perceived from different perspectives while two agents are acting and paying attention to specific aspects of the environment. Evaluation is extended to include how a robot can comprehend and produce characterizations of situational awareness and how it can use shared situation awareness to guide its own actions.

Assess with end-users. The integrated robot is assessed by end-users performing experiments with the new robot, the one from last year, and a tele-operated robot, in different situations and with task-varying time limits. The tasks focus on sharing situation awareness to guide actions from both robot and rescuer and the influence of this on both effectiveness as more psychological factors as emotional state and taskload.

Description of work – Milestones

The milestones from WP7 are the same as the integration and evaluation milestones for the entire project (§1.1.5 and §1.3.7): WP7 provides the integrated means to achieving the NIFTi objectives.

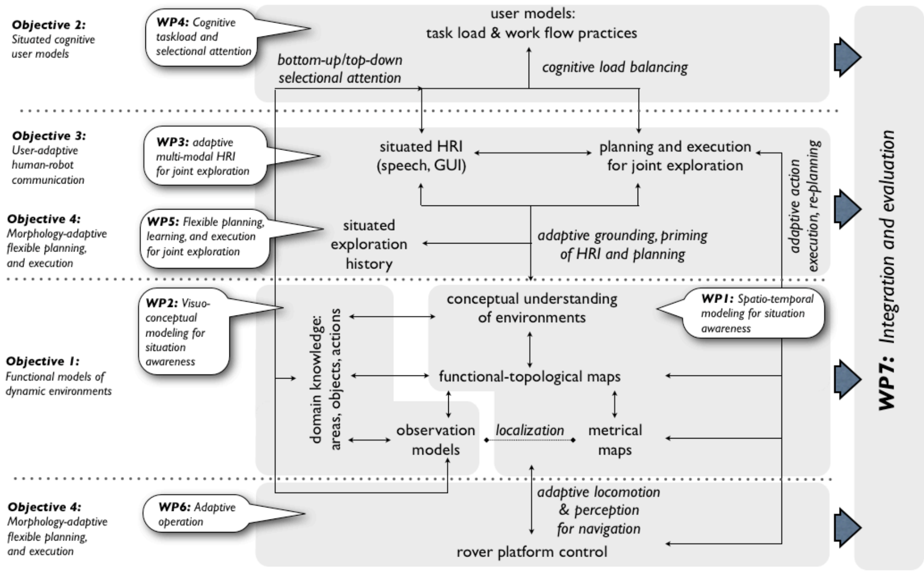
- Roadmap Yr1: Human-instructed exploration. Integrated cognitive robot system, end-user evaluation, and empirical domain analyses. (M12)
- Roadmap Yr2: Human-assisted exploration. Integrated cognitive robot system, end-user evaluation, and empirical domain analyses. (M24)
- Roadmap Yr3: In-field joint exploration planning. Integrated cognitive robot system, and end-user evaluation. (M36)

- Roadmap Yr4: Shared situation awareness. Integrated cognitive robot system, and end-user evaluation. (M48)

Description of work – Contributions to project

The diagram below depicts the integration interdependencies of WP7 with the functionalities of other WPs.

The novelties we bring to this project by the integration is the development of integrated, interactive robots for joint exploration using HRI. We establish this by applying user-centric engineering and evaluation practices for developing integrated cognitive systems and making use of context-sensitive real-time data storage, data access, and data segmentation for robot systems. It is in WP7 that NIFTi ultimately achieves putting the human factor into cognitive architectures.



Partner	Contributions to WP7 efforts
ALL	infrastructure – process communication protocols, real time data repositories and -access, logging, data "segmentation" (communicating partial models)
Fraunhofer	infrastructure, VM, coding guidelines, sensor evaluation, domain analysis (sensors, real time aspects, workflow), sensor integration (SW level)
BLUE	domain analysis (morphology), hardware integration
CTU	domain analysis (sensors, real time aspects); sensor integration (SW level)
DFKI	domain analysis (dialogue, spatial representations); integration of modalities and models for situated interaction history, and for symbol grounding; integration of modalities and models for hybrid maps; end user evaluation
ETHZ	domain analysis (dynamics, spatial representations, real time aspects), sensor integration (SW level), integration of modalities and models for hybrid maps
ROMA	domain analysis (tasks); end-user evaluation; integration of modalities and models for hybrid maps
TNO	domain analysis (work flow, use requirements, current use practices); end-user evaluation; human-centered design, cognitive engineering

Description of work – Risk management for WP7

Risk	Probability	Gravity	Contingency Plan
Communication between partners and with end users	medium	medium	Extra workshop and user training
Handling vast amounts of data	medium	medium	Using data base storage server (MySQL, SAP)
Complexity of the platforms and sensors	medium	medium	Divide et. impera, use of agile methods, permanent test-driven integration
Limited, noisy or erroneous hardware especially under extreme conditions	medium-high	medium	Direct evaluation with end users on system and sensor level, hardware adaptation / extension

Description of work – Evaluation metrics

Explicit targets are described with the individual tasks, above. Below we detail the different metrics for evaluating different aspects of the tasks.

Scenarios and use cases:

- Completeness and validity: Completeness will be evaluated by comparing it with task analyses and the validity by verification of the scenarios and use cases by domain experts (including end-users).

System integration: Metrics for assessing the performance of the integrated system are:

- Data processing time,
- Number of errors, (e.g. dialogue – speech recognition, comprehension errors; map visualization – wrong walls)
- Integration between sensors,
- Interaction between sensors and actors,
- The capability for morphology adaptation.

Assess technical capabilities: The robot will be tested every year against a tele-operated robot and the robot developed in the previous year. The robots will act in the NIST arena as described in the concerning tasks. For robot performance, we distinguish the following components:

- Time for specific actions,
- Navigation path,
- Detection of objects (e.g. victims, terrain identification, threats).

Assess with end-users: Each system is evaluated with end-users according to real world use cases, employing e.g. the NIST performance metrics for USAR environments [101] and field tests. In this task we will assess the overall joint Human-Robot performance (effectiveness, efficiency) and assess the human operator's behaviour to attune the dialogue strategies, operation modes and adaptation mechanisms:

- Effectiveness, (accuracy and completeness – victims found, threats avoided),
- Efficiency (required resources such as time (time spent to control robot, time spent to see whole area),
- Satisfaction (questionnaires e.g. [151]),

- Situation awareness (questionnaires e.g. [51] threats, noticed terrain features, noticed threats),
- Cooperation ([179]),
- HRI ([235], time spent to control, used dialogue),
- Cognitive task load (SOWAT tool [76]),
- Emotional state (questionnaires),
- Trust (questionnaire [237]).

Deliverables:

Yearly end-user evaluation reports describe the functionality for achieving the yearly WP milestones, their evaluation, and applicable domain analyses. The software prototypes provide a yearly complete integration of that functionality. Software is documented (API, user docs).

DR 7.1.1 *Specification of project software development and -documentation standards* (R; M3) [4PM]

DR 7.1.2 *Specification of current human factor knowledge of the USAR domain and known metrics and collaborative tools in this domain.* (R; M6) [3PM]

DR 7.1.3 *Integration and evaluation for human-instructed exploration.* (P, R; M12) [22.5PM]

DR 7.2.4 *Integration and evaluation for human-assisted exploration.* (P,R; M24) [31.5PM]

DR 7.3.5 *Integration and evaluation for in-field joint exploration planning.* (P,R; M36) [32.5PM]

DR 7.4.6 *Integration and evaluation for sharing situation awareness.* (P,R; M44) [33.5PM]

1.3.5.8 WP8: Dissemination and community building .

Work package number:	8	Starting date or starting event						1
Work package title	Dissemination and community building							
Activity type	OTH							
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	4	3	2	2	3	4	2	0	0

Objectives

WP8 provides effective means for NIFTi to disseminate its results as widely as possible, to maximize its impact. NIFTi adopts traditional means of publicizing project results in scientific journals and high-profile conferences (aiming for a high percentage of joint publications), and puts in place an efficient dissemination plan that combines several innovative means for dissemination (e.g. Rescue Days, scientific workshops, trade fairs, NIFTi project portal, public appearance) with community building (summer schools, community, presentation in research networks; cf. §2.4.5).

Description of work:

1. *Scientific publications*: Publications at high-level conferences and in high-impact journals, preferably joint publications. (M1-48)
2. *NIFTi portal*: Establish a web presence for the project, and for a larger community of interested third parties. Publicly accessible: publications, software, videos, newsletters, and evaluation testbeds. Intranet: VM distributions, internal reports, wikis for quality assurance guidelines, software development. (M1-48)
3. *Rescue Days*: Annual open days for rescue organizations, industry, academia, and the media. (M12-48)
4. *Trade fairs*: Present NIFTi at USAR-related trade fairs, once a year. (M12-48)
5. *Newsletter*: Bi-annual newsletter with latest project efforts and results, for relevant industrial and scientific contacts, and members of the USAR community. (M1-48)
6. *Summer schools*: Annual summer schools for 40 researchers from within and outside the project; tutorials and hands-on experience. (M1-48)
7. *Partner exchange program*: Program for extended working visits (2-3 weeks) of PhD-level staff at partner sites. (M1-48)
8. *Scientific workshops*: Annually organized workshop, in conjunction with scientific conference or at conference center. (M12-48)
9. *Public appearance*: Establishing active appearance in public media, through existing private and corporate communication contacts. (M1-48)

Deliverables:

DR 8.1.1 *NIFTi project portal.* (P; M3)

DR 8.1.2 *Market analysis for USAR robots with HRI.* See §3.2.1. (R; M10)

DR 8.1.3 *Proceedings of the NIFTi summer school Year 1.* (R; M12)

DR 8.2.4 *Proceedings of the NIFTi summer school Year 2.* (R; M24)

DR 8.3.5 *Proceedings of the NIFTi summer school Year 3.* (R; M36)

DR 8.4.6 *Updated market analysis for USAR robots with HRI.* See §3.2.1. (R; M44)

DR 8.4.7 *Proceedings of the NIFTi summer school Year 4.* (R; M48)

DR 8.4.8 *Public release of the open source NIFTi software.* (R; M48)

1.3.5.9 WP9: Management .

Work package number:	9	Starting date or starting event							1
Work package title	Management								
Activity type	MGT								
Participant number	1	2	3	4	5	6	7	8	9
Participant short name	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Person months	10	2	2	2	2	2	2	0.80	0.5

Objectives

The objective of this WP is to coordinate the administrative work within the consortium.

Description of work:

1. Providing input to meetings of the General assembly.
2. Preparing and conducting bi-annual meetings of the project's General Assembly.
3. Preparing annual progress reports for the Commission.
4. Preparing half-yearly summary briefings for the Project Officer.

Deliverables:

DR 9.1.1 *NIFTi annual progress report Year 1.* (R; M12)

DR 9.2.2 *NIFTi annual progress report Year 2.* (R; M24)

DR 9.3.3 *NIFTi annual progress report Year 3.* (R; M36)

DR 9.4.4 *NIFTi annual progress report Year 4.* (R; M48)

1.3.6 Efforts for the full duration of the project

Project Effort Form 1 – Indicative efforts per beneficiary per WP

Partic. no.	Partic. name	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	Total PMs
1	DFKI	8	12	50	20	10	0	20	4	10	134
2	TNO	0	0	12	55	4	0	20	3	2	96
3	Fraunhofer	22	16	9	3	9	20	25	2	2	108
4	BLUE	2	0	0	0	2	18	4	2	2	30
5	ETHZ	62	8	0	0	6	27	20	3	2	128
6	CTU	8	55	0	4	5	0	18	4	2	96
7	ROMA	6	9	19	8	42	8	14	2	2	110
8	FDDo	0.25	0	0.75	5.5	0	0.25	3	0	0.80	10.55
8	VVFF	0	0	1.5	6	7	0.25	3	0	0.5	18.25
	Total	108.25	100	92.25	101.5	85	73.5	127	20	23.30	730.80

Project Effort Form 2 – Indicative efforts per activity type per beneficiary

Activity type	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF	Total Activ.
RTD/Innovation										
WP1	8	0	22	2	62	8	6	0.25	0	108.25
WP2	12	0	16	0	8	55	9	0	0	100
WP3	50	12	9	0	0	0	19	0.75	1.5	92.25
WP4	20	55	3	0	0	4	8	5.5	6	101.5
WP5	10	4	9	2	6	5	42	0	7	85
WP6	0	0	20	18	27	0	8	0.25	0.25	73.5
WP7	20	20	25	4	20	18	14	3	3	127
Total 'research'	120	91	104	26	123	90	106	9.75	17.75	687.50
Demonstration activities										
Total 'demonstration'	0	0	0	0	0	0	0	0	0	0
Consortium management activities										
WP9	10	2	2	2	2	2	2	0.80	0.5	23.30
Total 'management'	10	2	2	2	2	2	2	0.80	0.5	23.30
Other activities										
WP8	4	3	2	2	3	4	2	0	0	20
Total 'other '	4	3	2	2	3	4	2	0	0	20
Total Beneficiaries	134	96	108	30	128	96	110	10.55	18.25	730.80

1.3.7 List of milestones

The milestones follow the phases on the development cycle. For each year, we first define a choice point milestone for specification of required functionality. This choice point is based on domain analyses, end-user requirements, and (for Yr2-Yr4) past end-user evaluations. Following up on the choice point milestone is the target milestone for each year, based on the use cases in the roadmap. Exact comparisons are made against the targets in the roadmap (system-level; cf. WP7) and the targets for the individual components (component-level; cf. WPs 1–6).

Mile-stone#	Milestone name	WPs	Lead benef. responsible for milestone	Delivery date from Annex I	Means of verification
M1	Functionality for human-instructed exploration	WPs 1–7	DFKI	6	Choice Point: functionality to adopt for achieving M2. Based on statistically significant domain analyses, end-user requirements, available state-of-the-art.
M2	Human-instructed exploration	WPs 1–7	Fraunhofer	12	Target comparison: comparison to system-wide targets (roadmap; WP7) and component-level targets (WPs 1–6)
M3	Functionality for human-assisted exploration	WPs 1–7	DFKI	18	Choice Point: functionality to adopt for achieving M4. Based on statistically significant domain analyses, end-user requirements, and end-user evaluation Yr1.
M4	Human-assisted exploration	WPs 1–7	Fraunhofer	24	Target comparison: comparison to system-wide targets (roadmap; WP7) and component-level targets (WPs 1–6)
M5	Functionality for in-field exploration	WPs 1–7	DFKI	30	Choice Point: functionality to adopt for achieving M6. Based on statistically significant domain analyses, end-user requirements, and end-user evaluation Yr1-2.
M6	In-field exploration	WPs 1–7	Fraunhofer	36	Target comparison: comparison to system-wide targets (roadmap; WP7) and component-level targets (WPs 1–6)
M7	Functionality for sharing situation awareness	WPs 1–7	DFKI	42	Choice Point: functionality to adopt for achieving M8. Based on statistically significant domain analyses, end-user requirements, and end-user evaluation Yr1-3.
M8	Sharing situation awareness	WPs 1–7	Fraunhofer	48	Target comparison: comparison to system-wide targets (roadmap; WP7) and component-level targets (WPs 1–6)

2 Implementation

2.1 Management structure and procedures

The organisational structures for NIFTi are aimed at ensuring competent project management, both for day-to-day issues, and relative to the long-term project goals. The main elements are listed below. The formal powers and duties of these are specified in the consortium agreement.

- The General Assembly (GA)
- The project coordinator (PC)
- The project administrator who runs the project office (PO)
- The Executive Board (EXB)
- The Scientific Advisory Board (SAB)

The General Assembly The General Assembly is the ultimate decision-making body of the consortium, and is composed of one duly authorised representative of each party with an equal voting right. The GA is responsible for matters related to the consortium agreement, budget allocation, and the general direction of the project. Ordinary and extraordinary meetings of the General Assembly shall constitute a quorum if more than two-thirds of the Parties are present or duly represented by a proxy. Ordinary meetings will be held in twice in each reporting period beginning with the project kick-off meeting.

The General Assembly chair (Geert-Jan Kruijff, DFKI) will preside over meetings of the GA and be responsible for compile the reports about technologies to integrate in each cycle.

Extraordinary meetings of the General Assembly can be convened by the project coordinator, or at the request of the majority of the partners or of the majority of the executive board.

The Project Coordinator and Project Office The project coordinator (PC; Geert-Jan Kruijff, DFKI) is the single point of contact between the European Commission (EC) and the Consortium. The PC is responsible for the overall management of the project. He chairs the Executive Board, and prepares the meetings and records the decisions of the General Assembly and the Board. The PC is supported in his duties by the project administrator in the Project Office (PO). The PO supports the PC in his day-to-day business and is responsible for daily administrative and organisational issues.

The Executive Board (EXB) The executive board (EXB) is the executive committee of the consortium. The EXB supports the PC in fulfilling obligations to the EC, preparing the agenda items for GA meetings, managing the project, supervising the scientific and technical progress in the project. Finally, the EXB is responsible for coordinating the various activities for education, training and dissemination. The EXB shall convene once a month, via a phone or video conference, to review progress on the basis of partner progress reports submitted the week before. The EXB is particularly important in ensuring progress in and completion of tasks within the work packages and the attainment of milestones and deadlines.

The Scientific Advisory Board (SAB) The Scientific Advisory Board (SAB) will advise the EXB, helping in identifying risks and further potentials of the research. The SAB will consist of three internationally respected senior scientists from research fields relevant to the project.

Voting procedures The voting procedures are described in detail in the consortium agreement. What follows is a summary of those procedures. (The exact formulation in the consortium agreement takes precedence over the summary provided here.)

- The quorum for the GA and EXB is two-thirds (2/3) or more of its Members.
- Each Member of a Consortium Body present or represented in the meeting has one vote.
- Defaulting Parties may not vote.
- Decisions are to be taken by a majority of two-thirds (2/3) of the votes.

A Party which can show that its own work, time for performance, costs, liabilities, intellectual property rights or other legitimate interests would be severely affected by a decision of the GA or EXB may exercise a veto with respect to the corresponding decision or relevant part of the decision. When the decision is foreseen on the original agenda, a Member may veto such a decision during the meeting only. When a decision has been taken on a new item added to the agenda before or during the meeting, a Member may veto such decision during the meeting and within 15 days after the draft minutes of the meeting are sent. In case of exercise of veto, the Members of the related Consortium Body (GA or EXB) shall make every effort to resolve the matter which occasioned the veto to the general satisfaction of all its Members.

For each meeting, the chairperson of that meeting will produce written minutes. These minutes constitute the formal record of all decisions taken at that meeting. These minutes are to be sent to all members within 14 days of the meeting.

2.2 Beneficiaries

Table 3 (p. 107) lists the level of involvement of the key project contributors in this project, and in other ongoing projects. Table 4 provides an overview of partner expertise, and how this contributes to achieving the objectives.

2.2.1 German Research Center for Artificial Intelligence (DFKI)

Organization DFKI (1988) is today one of the largest nonprofit contract research institutes in innovative technology based on Artificial Intelligence (AI). DFKI is based in Kaiserslautern, Saarbrücken, and Bremen. R&D is carried out in several research labs, including Language Technology (LT; Uszkoreit), Intelligent User Interfaces (IUI; Wahlster), and Robotics (RL; Kirchner). DFKI has close ties with the Universities of Kaiserslautern, Saarbrücken, and Bremen.

Contributions and experience DFKI contributes to multi-modal HRI for joint exploration (WP3), system integration (WP7), and connecting HRI, planning, user models and situation awareness (WPs1–5). Particularly to deal with time-critical HRI (WP3), DFKI builds on extensive experience in building (and analysing the performance of) fast robust methods for natural language processing (LT; fast and responsive systems for grammatically processing large-scale amounts of data), user- and context adaptive multi-modal dialogue systems for Question-Answering, mobile applications, HRI, and in-car applications (LT; IUI), and integrated systems for HRI (LT, particularly through EU FP6 IP "Cognitive Systems for Cognitive Assistants (CoSy)" and the EU FP7 "Cognitive Systems that Self-Understand and Self-Extend (CogX)").

Staff *Dr.ir. Geert-Jan M. Kruijff* (1970) is the NIFTi coordinator, at DFKI. He is a senior researcher/project leader in the LT Lab. He holds a PhD in computer science from Charles University, Prague (2001). His research focuses on developing cognitive architectures which model the dialogue capabilities of a robot, and connecting dialogue with models of a robot's experience. He has over 90 publications in the fields of human-robot interaction, spatial modeling, situated dialogue processing, and cognitive architectures. `http://www.dfki.de/~gj`

Dipl.-Ling. Hendrik Zender a researcher in the Language Technology Lab at the German Research Center for Artifical Intelligence (DFKI GmbH) and a PhD student at Saarland University.

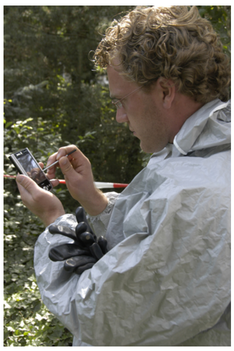
His research interests lie in linguistic aspects of spatial cognition for autonomous robots, and in human-robot interaction in general. `http://www.dfki.de/~zender`

Dr. Stephan Busemann (1957) is the associate head of DFKI's Language Technology Lab, where he is working as a Principal Researcher, lab manager and project leader. In 2000, he was appointed DFKI Research Fellow. His expertise are Artificial Intelligence, Computational Linguistics, Language Technology, and Natural Language Generation. Present research interests focus on applied natural language processing, including natural language generation. `http://www.dfki.de/~busemann`

- Kruijff, G.J.M., Lison, P., Benjamin, T., Jacobsson, H. Zender, H., Kruijff-Korbayová, I. "Situated dialogue processing for HRI." In H. Christensen, G.J.M. Kruijff, J. Wyatt (eds.) *Cognitive Systems*. Springer. 2009.
- Zender, H., Martínez Mozos, O., Jensfelt, P., Kruijff, G.J.M., Burgard, W. "Conceptual Spatial Representations for Indoor Mobile Robots." Robotics and Autonomous Systems. 56(6):493-502, June 2008.
- Jacobsson, H., Hawes, N., Kruijff, G.J.M., Wyatt, J. "Crossmodal Content Binding in Information-Processing Architectures." Proc. of the 3rd ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI). March 2008.
- Lison, P. and Kruijff, G.J.M. "An integrated approach to robust processing of situated spoken dialogue." In Proc. Intl. Wsh. on Semantic Representation of Spoken Language (SRSL'09). April 2009
- Hawes, N. Sloman, A.,Wyatt, J.W., Zillich, M., Jacobsson, H., Kruijff, G.J.M., Brenner, M., Berginc, G. and Skocaj, D. "Towards an integrated robot with multiple cognitive functions." In Proceedings of the 22nd Conference on Artificial Intelligence (AAAI-07), 2007.
- Kruijff, G.J.M., Zender, H., Jensfelt, P. and Christensen, H.I. "Situated Dialogue and Spatial Organization: What, Where... and Why?" International Journal of Advanced Robotic Systems. Vol. 4, No. 2. 2007

2.2.2 Netherlands Organization for Applied Scientific Research (TNO)

Organization TNO is an independent, not-for-profit, body with a public mission. TNO Human Factors (HF), is one of the main human factors laboratories worldwide and has 200 employees. Its mission is to optimize human performance in complex and demanding environments. Its multidisciplinary staff includes psychologists, interaction designers, engineers, and physicists. It has extensive, advanced tools and facilities for human-in-the-loop testing, e.g. emergency rooms, virtual reality. Furthermore it has good contacts with Dutch firefighters, police, defense, and the Dutch USAR organisation.



Contributions and experience TNO HF contributes primarily to aspects of adaptive interfaces for decreasing the workload of the user, supporting the user in decision processes (WP3), improving the situation awareness of user and device (WP4), and establishing "natural" human-robot interaction (WP3) with the help of cognitive models, and affect recognition and expression. The adaptive interface will be designed and evaluated according to a human-centered, cognitive engineering approach (WP6) [179]. This knowledge is applied in international and national projects. TNO HF leads e.g. the MECA consortium wherein the aim is to empower cognitive capacities of human-machine exploration teams (inluding robots)[181] for long duration space missions. Other related projects are ePArtner wherein a robot is used to react affectively on the the user [150, 176], MultimediaN wherein both personalized adaptive interfaces are investigated [238], and the IOP-MMI project collaboration at a distance: supporting urban search & rescue missions in human-human teams. Furthermore, TNO HF has extensive experience with teleoperation and telepresence operation of UAVs and UGVs, thus human-robot teams [261, 105]. In this project our knowledge of the urban search and rescue task will be combined with our knowledge on UGV control human factor issues.

Staff *Prof. Dr. Mark A. Neerincx* (1960) is professor in Man-Machine Interaction at the Delft University of Technology, Department of Mediamatics, Faculty of Electrical Engineering, Mathematics and Computer Science. Furthermore, he is head of the Intelligent Interface group at TNO Human Factors. He has extensive experience in applied and fundamental research, within national and international consortia (often as project leader). Important results are (1) a cognitive task load model as foundation of a cognitive engineering method for task allocation and design of cognitive support, (2) models for adapting assistance to the individual user and momentary usage context. He is involved in the organisation of conferences, workshops and tutorials to discuss and disseminate human-factors knowledge.

Rosemarijn Looije (1982) received her Masters Degree in Artificial Intelligence/Man Machine Interaction in 2006 from the University of Groningen. She is now a research member of the Intelligent Interface group at TNO HF. Her master thesis involved the persuasion of chronically ill to adhere to their treatment by using a robot that reacted affective on the affective state of the user.

Chris Jansen (1972) is a researcher/project leader at TNO in the Business Unit Human Factors. He has a background in cognitive/experimental psychology and human movement science, supplemented by education in mechanical engineering. His research and development projects concentrate on designing, building and evaluating man-machine interfaces for optimizing human performance in complex environments. Areas of interest are aircraft cockpit design, man-robot interactions, tactile displays, and mobile information and communication systems for dismounted soldiers and commanders. He holds a project management degree at level C of the International Project Management Association.

Dr. Jurriaan van Diggelen (1976) obtained a PhD degree in Artificial Intelligence for his thesis on heterogeneous ontologies in multi-agent systems. In 2007 and 2008, he has been a Postdoc working on adaptive information support for crisis management. He is now employed as a scientific researcher at TNO HF, where he is involved in several projects on Human-Agent Teamwork.

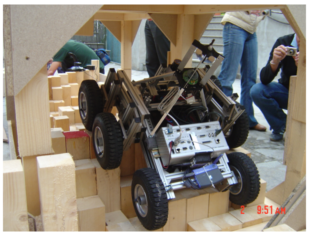
- M.A. Neerincx and J. Lindenberg (2008). Situated cognitive engineering for complex task environments. In: Schraagen, J.M.C., Militello, L., Ormerod, T., and Lipshitz, R. (Eds). Naturalistic Decision Making and Macrocognition (pp. 373-390). Aldershot, UK: Ashgate Publishing Limited
- Neerincx, M.A. (2003). Cognitive task load design: model, methods and examples. In: E. Hollnagel (ed.), Handbook of Cognitive Task Design. Chapter 13 (pp. 283-305). Mahwah, NJ: Lawrence Erlbaum Associates.
- R. Looije, M.A. Neerincx, and G.J.M. Kruijff. "Affective Collaborative Robots for Safety & Crisis Management in the Field." Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2007). May 2007. Delft, The Netherlands.
- J. van Diggelen, R.J. Beun, R.M. van Eijk, P.J. Werkhoven, Agent Communication in Ubiquitous Computing: the Ubismart Approach, Proceedings of the Seventh International Conference on Autonomous Agents and Multi-agent Systems (AAMAS08), ACM Press, pp. 813-820, 2008

2.2.3 Fraunhofer Institut für Intelligente Analyse- u. Informationssysteme (FhG-IAIS)

Organization The FhG-IAIS is an institute of the Fraunhofer Society, Europe's largest research organization. FhG-IAIS now has about 270 employees, including a research staff of about 170. FhG-IAIS focuses on research and development on innovative systems for data analysis and information extraction, in software and in hardware. Its main business areas include Information Mining, Geo Intelligence, Media Information Systems, Exploration Robotics, and Educational Robotics. In these areas, FhG-IAIS performs basic and pre-market research leading to application-oriented concepts and individual solutions for industrial, scientific, and governmental clients. FhG-IAIS is closely cooperating with universities from its home region, particularly with the University of Bonn and the University for Applied Sciences Bonn-Rhein-Sieg in Sankt Augustin.

Contributions and experience The business area most related to the project is 'Exploration Robotics'. The fields of application include security technology, rescue missions, measurement and monitoring of buildings [189, 240, 190, 192, 193, 66, 191, 239]. For realising complete robotics systems, FhG-IAIS possesses the required competences in robot control architectures, sensorics, and system integration (WP1,WP2,WP7). The FhG-IAIS lab facilities include a robotics workshop for constructing and building robot prototypes and their mechatronics components(WP 6), an outdoor experimental facility related to sewer robotics activities and other labs.

FhG-IAIS has also experiences in real disaster since a robotic team under guidance of Dr. Surmann helps together with Prof. Robin Murphy and Prof. Satoshi Tadokoro at the building collapse in cologne in March 2009. FhG-IAIS is active in the RoboCup federation (www.robocup.org) [190] and has excellent international contacts for promoting new developments, like the expected results of the NIFTi project. Fraunhofer IAIS disseminates the expected results of NIFTi into its organizational, national and international networks, and support the transfer of these results into industry and applications. Furthermore, FhG-IAIS participates in the EU funded projects OASIS, IRRIIS, SHARE, MACS and others. In the OASIS project, FhG-IAIS is responsible for developing a decision support system for crisis management.



Staff *Prof. Dr.-Ing. Hartmut Surmann* (1963) is a senior researcher and project leader within the FhG-IAIS. He received his diploma in Computer Science and his PhD in Electrical Engineering from the University of Dortmund, Germany, in 1989 and 1994, respectively. Since 2009 he is also Professor for Autonmous Systems at the Applied University of Gelsenkirchen. He received several awards, e.g., Vize World champion in RoboCup rescue 2004, the FUZZ-IEEE/IFES'95 robot intelligence award, NC2001 best presentation award, SSRR 2005 best paper award and the Ph.D. award for his thesis from the German AI institutes in 1996. In March 2009 he managed the international robot mission at cologne collapsed building directly at the disaster.

2.2.3.1 Staff *Dipl.-Inform. Rainer Worst* (1955) is a senior researcher within the FhG-IAIS. He received a diploma in Computer Science (Univ. Bonn, 1981) and started thereafter as an IT-Expert at a consulting company in Wiesbaden. In 1984 he joined the former GMD (National Research Center for Information Technology) in Sankt Augustin, which merged 1999 with Fraunhofer. Since 1984, he worked as a consultant, researcher and project leader in the areas Software Engineering, Quality Management, Autonomous Systems, and Virtual Environments. Recently, he participated in the EU projects MACS and INT-MANUS.

- H. Surmann, D. Holz, S. Blumenthal, T. Linder, P. Molitor, V. Tretyakov, "Teleoperated Visual Inspection and Surveillance with Unmanned Ground and Aerial Vehicles", iJOE, Vol. 4, No. 4, Nov. 2008. pages 26-38.
- A. Nüchter, K. Lingemann, J. Hertzberg, H. Surmann, K. Pervölz, M. Hennig, K.R. Tiruchinapalli, R. Worst, T. Christaller, "Mapping in rescue environments with Kurt3D (Best paper awarded after VizeWorld Champoins 2004)", Proc. of the SSRR 2005, pp. 158-163.
- T. Wisspeintner, A. Bose, P. Plöger, "Robot Prototyping for Rough Terrain Applications and High Mobility with VolksBot RT", Proc. SSRR 2006.
- A. Nüchter, K. Lingemann and J. Hertzberg and H. Surmann,"6D SLAM – 3D mapping outdoor environments", Journal of Field Robotics, Vol. 24, no. 8-9, pp. 699-722, 2007.
- S. Frintrop, E. Rome, A. Nüchter, H. Surmann, "A Bimodal Laser-Based Attention System", Journal of Computer Vision and Image Understanding (CVIU), Vol. 100, no. 1-2, pp. 124-151, 2005.
- A. Nüchter, K. Lingemann, J. Hertzberg, O. Wulf, B. Wagner, H. Surmann, "3D mapping with semantic knowledge", RoboCup International Symposium 2005, pp. 335-346.

2.2.4 BlueBotics (BLUE)

Organization BlueBotics SA is a Swiss SME located in Lausanne. It is a spin-off of the Swiss Federal Institute of Technology Lausanne (EPFL) and it has been founded in 2001 with the mission to market innovative and promising mobile robotics technologies developed at the Autonomous Systems Lab (today with the Swiss Federal Institute of Technology Zurich, ETHZ). The company started with the Shrimp outdoor platform and with the production and exploitation of eleven RoboX (tour-guide robots) for the Robotics exposition at Expo.02 - Swiss National Exhibition in Neuchâtel. BlueBotics's domain of expertise covers the engineering of mechatronic systems. Having experience in integration, design, prototyping, and production, the company can offer complete professional solutions or enhance current products with its technologies. Its main products and services include autonomous navigation technology, and service robotics.

Contributions and experience BlueBotics has a large experience in designing, building and commercializing rovers for rough terrain. The company started with the Shrimp outdoor platform (Swiss Technology Award 2001), which is today commercialized as research platform. The Shrimp has been scaled up to the SOLERO ESA (European Space Agency) rover. BlueBotics also participated to the ESA Exomars phase A by designing all the small scale prototypes (RCL-C, RCL-E and CRAB) and both the steering and hub drive units of the breadboard rover. Furthermore, BlueBotics is also involved as consultant for research institutes and co-authored several publications [245, 246, 249].

BlueBotics will be in charge of WP5 Adaptive locomotion and collaborate to WP1 Spatio-temporal modeling for situation awareness. This WP requires an important interaction with especially WP1, where BlueBotics will also collaborate to ensure a good interaction and integration of the developments. BlueBotics has participated to the BIBA EU project (IST-2001-32115) as subcontractor for the design and production of three research platforms. Since year 2002, BlueBotics is partner of the MOVEMENT STREP project (IST-2002-2.3.2.10), where the consortium aims to the development of a modular semi-autonomous / autonomous wheelchair system for disabled persons. Since January 2006, the company is also participating to the BACS IST-4-027140 IP project. Finally, on May 1, 2007, BlueBotics will start the robots@home FP6-2006-IST-6-045350 STREP project. BlueBotics received the following awards lately: 2007 - IEEE/IFR Invention and Entrepreneurship, Finalist Award, 2007 - CTI Startup Label - "ready for sustainable business development" and 2008 - EURON/EUnited Robotics Technology Transfer Award, 2. Prize.

Staff *Dr. Nicola Tomatis* (1973) received his M.Sc. in computer science in 1998 from the Swiss Federal Institute of Technology (ETH) Zurich. After working as assistant for the Institute of Robotics, ETH, he moved to the Swiss Federal Institute of Technology (EPFL) Lausanne, where he received his Ph.D. in 2001. His research covered metric and topological (hybrid) mobile robot navigation, computer vision and sensor data fusion. From 2001 to 2005 he was (part-time) senior researcher with the Autonomous Systems Lab, EPFL (now ETH), where he led several projects and continued his research in navigation, man-machine interaction and robot safety and reliability. He has authored/co-authored more than 35 journals and conference papers. In 2001 he joined BlueBotics SA. Since year 2003 he is CEO of the company.

Dr. Pierre Lamon (1974) received his M.Sc. in micro-engineering in 2000 and his Ph.D. in 2005 from the EPF, Lausanne. He received the Georges Giralt PhD Award for the best PhD thesis in Europe for his work titled "3D Position Tracking for all-terrain Robots". During his Ph.D. Pierre has participated to several ESA projects (such as SOLERO, EXOMARS). In 2005, he has worked as postdoc at the LAAS-CNRS, France to develop a decentralized data fusion with asequent and delayed data for multi-robot mapping and cooperation. In 2006, he was the technical project leader of the SmartTer project (EPFL): an autonomous Smart car for mapping of cities. He has authored/co-authored more than 15 journals and conference papers. Pierre joined BlueBotics at the end of 2006 as scientific manager and R&D member.

Grégoire Terrien (1976) received his M.Sc. in micro-engineering in 2001 from the EPF, Lausanne. Since 1999, he has been involved in the design and system integration of various robots and has leaded several robot-design projects for the Autonomous Systems Lab, EPFL. Grégoire joined BlueBotics in 2001 as R&D manager. He is today in charge of the system design and integration of all the new robots developed by BlueBotics. Grégoire has designed more than 20 different mobile robots.

2.2.5 Eidgenössische Technische Hochschule Zürich (ETHZ)

Organization The **Autonomous Systems Lab** (ASL) at the Eidgenössische Technische Hochschule Zürich ETHZ (until June 2006 at EPFL Lausanne, `http://asl.epfl.ch`) is an internationally renowned research lab in the field of autonomous robot design and navigation. It has a large experience in the design and autonomous navigation of wheeled and flying autonomous robots for different kinds of environments. Among recent results are personal robots with multi-modal interaction capabilities, wheeled locomotion systems that passively adapt to rough terrain [246, 52], autonomous micro-aircrafts [21], and autonomous cars with 3D navigation and mapping capabilities in rough terrains. Apart systems design, a major research focuses are in cognitive maps [264], feature based SLAM [152, 184] using multiple modalities and path planning in highly dynamic environments [122]. The ASL consists of around 45 researchers and application engineers and is involved in various National, European and ESA projects.

Contributions and experience The expertise relevant to NIFTi project is the Lab's mapping experience and especially the large competence in SLAM, hybrid and hierarchical representation of the environment [255, 254, 168], navigation in dynamic environments [231], and semantic and cognitive mapping [262, 264, 263]. The ASL team was responsible of the development and operation of 11 mobile tour-guide robots during the Swiss national exhibition expo.02. The Lab is also involved in the ESA ExoMars project, in inspection robotics for power plant applications, and autonomous cars. ASL is leading the EU projects on Bayesian Approaches to Cognitive Systems (BACS, `http://www.bacs.ethz.ch`) and on Fully Autonomous Micro-Helicopter (muFly, `http://www.mufly.ethz.ch`). The central commitment of ETHZ to NIFTi consists in the development and implementation of spatio-temporal cognition for situation awareness (WP1-leader). This involves novel 3D mapping and scene analysis technologies as well as navigation in cluttered environments. We are also strongly involved in robot architectures for variable morphology (WP6) and in integration and evaluation (WP7).

Staff *Prof. Dr. Roland Siegwart* (1959) is full professor for autonomous systems at ETH Zürich since July 2006. He has a Diploma in Mechanical Engineering (1983) and a PhD in Mechatronics (1989) from ETH Zürich. In 1989/90 he spent one year as postdoctoral fellow at Stanford University. After that he worked part time as R&D director at MECOS Traxler AG and as lecturer and deputy head at the Institute of Robotics, ETH Zürich. In 1996 he was appointed as associate and later full professor for autonomous micro-systems and robots at the Ecole Polytechnique Fédérale de Lausanne (EPFL). During his period at EPFL he was co-initiator and founding Chairman of Space Center EPFL and Vice Dean of the School of Engineering. In 2005 he held a visiting position at NASA Ames and Stanford University. Roland Siegwart is member of the Swiss Academy of Engineering Sciences and board member of the European Network of Robotics (EURON). He served as Vice President for Technical Activities (2004/05) and is currently Distinguished Lecturer (2006/07) and AdCom Member (2007-2009) of the IEEE Robotics and Automation Society. He is member of the "Bewilligungsausschuss Exzellenzinitiative" of the "Deutsche Forschungsgemeinschaft (DFG)" and coordinator of two EU projects.

Dr. Cédric Pradalier (1978) received the Diplôme d'Ingénieur from the École Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble (ENSIMAG), France, in 2001, and the Diplôme d'Études Approfondies (Master) from the Institut National Polytechnique de Grenoble (INPG), France, also in 2001. He received a PhD degree in Robotics from the INPG, in 2004. Subsequently, he joined the CSIRO ICT Centre, Brisbane, Australia as a Post-Doctoral Fellow in the Autonomous System Lab, from 2004 to 2007. In November 2007, he joined the ASL at ETH Zurich to take the position of Deputy Director.

Dr. Francis Colas (1979) is postdoctoral fellow at ASL since january 2009. He received a PhD degree in Computer Science from the INPG, in 2006. Subsequently, he joined the College de France, Paris, France as a Post-Doctoral Fellow in the Laboratoire de Physiologie de la Perception et de l'Action, from 2007 to 2008. In January 2009, he joined the Autonomous Systems Lab. His research interests include Bayesian modelling of perception and action applied from cognitive sciences to robotics.

- Ó. Martínez Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard. Supervised semantic labeling of places using information extracted from sensor data. Robotics and Autonomous Systems, 55(5):391-402, 2007.
- R. Triebel, P. Pfaff, and W. Burgard. Multi-level surface maps for outdoor terrain mapping and loop closing. In Proc. of the International Conference on Intelligent Robots and Systems (IROS), 2006.
- V. Nguyen and R. Siegwart. Information relative map: Going toward constant time slam. In Proc. of The European Robotics Symposium (EUROS), 2008.
- S. Vasudevan, S. Gaechter, V. Nguyen, and R. Siegwart. Cognitive maps for mobile robots: an object based approach. Robotics and Autonomous Systems, 55(5):59-371, 2007.
- T. Thueer, P. Lamon, A. Krebs, and R. Siegwart. Crab: Exploration rover with advanced obstacle negotiation capabilities. In Proc. of The 9th ESA Workshop on Advanced Space Technologies for Robotics (ASTRA), 2006.

2.2.6 Czech Technical University Prague (CTU)

Organization The *Center for Machine Perception* of the Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague is a 24 staff and 10 full-time PhD students strong research and educational unit (`http://cmp.felk.cvut.cz`). The center established in 1996 has become a renewed research lab in computer vision and pattern recognition. The center is currently involved in four EC funded research projects (DIRAC, e-Trims, DIPLECS, PRoVisG and two Marie-Curie Research Training Networks VISIONTRAIN and WARTHE) and in several national projects. Some researchers are paid from industrial grants, e.g., from Hitachi, Toyota Motor Company, Honeywell. Researchers of the center established two start-up companies – Neovision (`http://www.neovision.cz`) and Eyedea Recognition (`http://www.eyedea.cz`). The center keeps cooperating with both companies.

Contributions and experience The expertise relevant to NIFTi project is the computer vision expertise in general. In particular, four competencies are directly relevant to NIFTi project: 3D scene reconstruction from uncalibrated 2D views; Omni-directional cameras and its use for 3D vision tasks; Robot localization based on view-based representations; and Detection of locally described object which were learned in the supervised manner; functional environment models, particularly classifying and representing affordances of local visual objects (experience gained in EC project ActIPret).

The central role and commitment of CTU to NIFTi consists of design and implementation of computer vision-based modules allowing the robot to perceive the surrounded world. CTU leads the WP2 – Visuo-conceptual modeling for situation awareness. CTU will provide to the project needed computer vision expertise, will help to integrate other perceptual modalities.

Staff *Prof. Václav Hlaváč* (1956) has been a full professor of cybernetics at the Czech Technical University since 2002, `http://cmp.felk.cvut.cz/~hlavac`. His MSc. (1981) is in electrical engineering and PhD (1987) in cybernetics from the same institution. He was a postdoctoral fellow at the University of Sussex in Brighton, UK for 10 months in 1989, at the Technische

Universität Wien, Austria for 6 months in 1995 and in the Hitachi Central Research Laboratory in Tokyo, Japan for six months. Prof. Hlaváč's expertise is in 3D computer vision [225], 3D reconstruction, omni-directional vision, motion analysis from video [247] and statistical pattern recognition [220]. Besides others, he is a co-author of the advanced textbook [230] which have been first or second in sales in US market since its publishing.

Dr. Tomas Pajdla (1969), assistant professor at CTU since 1995, male, M.S. and Ph.D. in EE at the Czech Technical University Prague, `http://cmp.felk.cvut.cz/~pajdla`. T. Pajdla has experience in geometry and algebra of computer vision, visual robot control, image matching, eye-hand calibration and coordination, precise digital optical measurements, photogrammetry, robot navigation using vision, image matching and object recognition. He published more than 50 scientific works in scientific journals and refereed conferences. In 2005, his students ranked second in the ICCV camera localization contest. T. Pajdla will mainly contribute to WP2, in particular to context map building and to visual context recognition. His publications relevant to the project are, e.g., [243], [158], [154], [84]. He has lead CTU team in the EC funded project PRoVisG - Planetary Robotic Vision Ground Processing (2008-2011).

Dr. Tomas Svoboda (1972) received his PhD degree in artificial intelligence and biocybernetics from the Czech Technical University in Prague in 2000. He spent three years as a post-doc with the Computer Vision Group at the ETH Zurich (Swiss Federal Institute of Technology) in the group of Prof. Luc van Gool. He is currently senior research fellow at the Czech Technical University. He has published on multicamera systems, omnidirectional cameras, image based retrieval, and learnable tracking methods. the publications have around 300 citations in the ISI Web of Knowledge (Thomson-Reuters) database with h-index 10. His current research interests include multicamera systems for scene understanding and telepresence, tracking and motion analysis. His public domain SW for multicamera calibration is widely used. His publications relevant to the project are [243], [242], [281].

Dr. Karel Zimmermann (1978), researcher, male, M.S. and Ph.D. in EE at the Czech Technical University Prague, `http://cmp.felk.cvut.cz/~zimmerk`. His research focus is in image sequences analysis and object tracking. He was awarded 'Antonin Svoboda Award' for the best informatics PhD thesis in the Czech Republic by the Czech Society of Cybernetics and Informatics. His recent publication relevant to the project is [281].

2.2.7 Sapienza, University of Roma (ROMA)

Organization The "Dipartimento di Informatica e Sistemistica" at the "Sapienza", University of Roma, was established in 1983 for the development of advanced research, innovative applications, and professional skills in information technology, automation and control, operations research and management. The AI group is one of the biggest AI group in the world, leading research in Cognitive Robotics, Vision and Perception, Learning, Reasoning about Actions, Knowledge Representation, Constraint-based Architectures for Planning and Scheduling, Description Logic, and Multi-robot and multi-agent systems. ALCOR (Autonomous Agent Laboratory for Cognitive Robotics) founded in 1998 is composed of nine members (assistant researchers, PhD students, collaborators and two full professors). The team works on the analysis, design and development of control systems for autonomous agents integrating perception, reasoning, and learning. ALCOR team participates since 2002 to the real-robot rescue competitions organized by NIST in Robocup. In Robocup-2004 ALCOR obtained the third-place cup, mentioning the high performance in visual perception. The team is currently involved in the FP6 EU-funded project VIEWFINDER, EU-project MAGICSTER, and other Italian projects with ASI, and MIUR projects FIRB ASTRO and a COFIN.

Contributions to NIFTi ALCOR contributes primarily to cognitive execution, skill learning, developing biologically inspired methods for attentive bottom-up and task-driven exploration of dynamic environments and flexible behaviours modeling (WP5), and also (WP3-4). Furthermore

ALCOR collaborates in the context of compositional structuring of functional representations (WP1) and visual processing (WP2), concentrating on multiple view data association, and object recognition for situation awareness. ALCOR team is also involved in the integration and evaluation efforts (WP6-WP7).

Alcor team has developped formal models for reasoning about action and perception and has extensive experience in the design and development of model-based systems especially in rescue scenarios. Has participated to Robocup Rescue competitions and organized events with the Civil Protection, the Fire department and other national security organizations. Great experience on unknown and hostile environment has been also acquired in ASI-funded projects.

Staff The key research scientist of the project is *Fiora Pirri*. She obtained the PhD at Université Pierre et Marie Curie, Paris 6, and is currently full professor of Computer Science with the DIS, University of Roma "La Sapienza". Her main research is in Cognitive Robotics, and in particular in perception and reasoning. In 1998 she founded the Autonomous Laboratory for COgnitive Robotics (ALCOR). She is member of AAAI, ACM and IEEE, she is a founding member of AI*IA, and permanent member of the Cognitive Robotics steering committee.

- Carbone, A., Finzi, A., Orlandini, A., and Pirri, F.. "Model-based control architecture for attentive robots in rescue scenarios." Autonomous Robots 24, 1:87–120, 2008.
- Belardinelli A., Pirri F., and Carbone A. " Motion saliency maps from spatiotemporal filtering". In "Attention In Cognitive Systems", Paletta and Tsotsos Editors, volume LNAI 5395, pages 112–123. Springer, NLD, 2009.
- Belardinelli A., Pirri F., and Carbone A. "Gaze motion clustering in scan path estimation." COGNITIVE PROCESSING, 9:269–282, 2008.
- Marchegiani L., Pirri F., and Pizzoli M. "Multimodal speaker recognition in a conversation scenario". In International Conference on Computer Vision Systems(ICVS), volume LNCS, September 2009. Springer.
- Marra S. and Pirri F. "Eyes and cameras calibration for 3d world gaze detection". In International Conference on Computer Vision Systems (ICVS), volume LNCS, pages 216 –227, NLD, May 2008. Springer.
- Lesperance Y., Lakemeyer G., Peters J., and Pirri F. "Proceedings of the 6th International Cognitive Robotics Workshop". IOS press, 1-91, 2008.

Professor *Marco Schaerf*, at the University of Roma 'La Sapienza', is currently the coordinator of the Rome unit of the European project MagiCster. Present president of the Italian Artificial Intelligence Association (AI*IA), his research interests include theory of actions and AI-applications of computer graphics. He received the excellence award from AI*IA in 1996.

- M. Fratarcangeli, M. Schaerf, R. Forchheimer: "Facial motion cloning with radial basis functions in MPEG-4 FBA ". Graphical Models 69(2): 106-118 (2007).
- Marco Cadoli, Marco Schaerf: "Partial Solutions with Unique Completion". Reasoning, Action and Interaction in AI Theories and Systems 2006: 101-115.

2.2.8 End-user organizations with beneficiary status: FDDo, VVFF

NIFTi involves two end-user organizations as beneficiaries: the Research Institute for Fire Service and Rescue Technology, of the Fire Department of Dortmund (FDDo), and the Corpo Nazionale Vigili del Fuoco, of the Italian Ministery of Interior (VVFF). These two beneficiaries are described in more detail in §2.3.1 below, together with the other end-user organizations that NIFTi involves as subcontractors. For the FDDo description see §2.3.1.9 (p.103), and §2.3.1.10 for VVFF (p.103)

2.3 Consortium as a whole

The consortium has been selected on the basis of the excellence of the research track record of each participant. The participants bring the necessary experience into the project to address the project objectives, realize them, evaluate them together with end-users, and establish possibilities for commercialization. The matrix in Table 5 indicates the required competences organized per each objective. This matrix complements Table 4 (p.108) which lists the specific expertise that is being contributed by the individual partners.

Where there are partners who have expertise in the same broad area we have carefully ensured that their expertise within these areas is mutually complementary. ETHZ, BLUE and Fraunhofer have expertise in (rescue) field robots with different morphology types. DFKI and ROMA both have experience in developing cognitive architectures, with ROMA focusing on logical planning-based approaches, and CTU and DFKI dealing with the issues in connecting representations across different modalities ("symbol grounding"). In spatial modeling, Fraunhofer and ETHZ have developed complementary methods to 3D-mapping of outdoor environments, whereas DFKI has focused on developing (multi-layered) qualitative spatial representations and their connections to reasoning, planning and dialogue. In perception, CTU has focused primarily on vision, ROMA on other types of sensory information relevant to rescue missions, and Fraunhofer has substantial experience in sensor modeling and -fusion. DFKI and TNO have investigated different aspects of human-robot interaction, DFKI focusing on spoken dialogue in HRI whereas HRI contributes experience in evaluating HRI. As for cognitive models, DFKI contributes models for linguistic processing load, whereas TNO has experience in cognitive user modeling, including task load. Table 6 summarizes the above.

There are several key features of this consortium that ensure that integration and communication will be outstanding in NIFTi. First of all we have selected several sites which offer multiple areas of expertise, notably ETHZ, Fraunhofer, ROMA, and DFKI. This has helped to keep the consortium to a manageable size, and this in turn means that the difficulties of integration and communication typically encountered in much larger consortia are lessened.

Most of the partners have been involved in numerous projects conducted in the FP6 and FP7 "Cognitive Systems" programs, as described in the participant profiles. A significant subset of the partners are thus well aware of the issues involved in developing integrated cognitive robots. Furthermore, several of the partners have worked together on previous successful projects, and are thus intimately familiar with each other's approaches and working methods.

All the partners have significant experience of working in, and managing large research projects with multiple partners, so that there is a wealth of experience to draw on in project management.

Finally, we also have the essential experience in building the large integrated cognitive robotic systems of the type we will investigate in WP7. DFKI, Fraunhofer, ROMA, and ETHZ all have experience in delivering large working systems that combine multiple complex sub-systems. Thus we understand the scientific issues, as well as the practical concerns in producing experimental platforms that push the state of the art and are technically sound.

2.3.1 End-user organizations

NIFTi wants to make a genuine contribution to ("impact on") USAR. NIFTi therefore considers collaboration with potential end-users, i.e. fire departments and rescue organizations, to be necessary. NIFTi involves several organizations with established experience in USAR missions: The IFR / Fire Department of Dortmund (FDDo, Germany), the Corpo Nazionale Vigili del Fuoco (VVFF, Italy), the Einsatzkommando Katastrophenhilfe Bereitschaftsverband (EiKdo, Switzerland), and RUAG Land Systems (RUAG, Switzerland). The organizations have been selected on the basis of their varied experience in USAR, as well as experience in specifying requirements for USAR technology, and evaluating such technology. Furthermore, several organizations (FDDo, VVFF, EiKdo) provide NIFTi access to their training areas. These areas are geographically "optimally" distributed to provide easy access to partners who need access to them (Germany, Switzerland, Italy).

End user organizations are involved in four different types of activities:

- **Data collection**: creation of data sets (image, video, sensor signals) based on observations in a training area
- **Knowledge transfer**: transfer of domain knowledge from experts to NIFTi, with the purpose of using this knowledge to support data interpretation and domain model building

- **Specification**: specification of requirements for platform functionality, and system functionality for NIFTi roadmap scenarios.
- **Evaluation**: user evaluation of components and systems developed within NIFTi.

All these activities are limited in time and scope. Their focus is to provide input to the NIFTi research. Domain knowledge transfer and evaluation activities are considered core to NIFTi and are therefore carried out by end user organizations with beneficiary status. End user organizations with small involvement in specification and data collection activities are subcontracted for. Table 7 (p.110) describes the involvement of the end user organizations in tasks (deliverables) for WPs 3–6. Detailed descriptions of these tasks are provided below. Table 8 (p.111) details the involvement of the end users in the specification- and evaluation phases in WP7. We refer to the WP7 description (§1.3.5.7) for descriptions of these tasks.

The two end user organizations involved as beneficiaries in NIFTi (FDDo, VVFF) contribute additional person months, over those being financially budgeted for. The person months mentioned in Table 7 and Table 8, as well as those in the WP summaries, regard the *total* number of person months (financed and contributed). The two subcontracted end user organizations (EiKdo, RUAG) budget for travel costs and training area use.

2.3.1.1 Involvement: Sensor-profiling in training areas (WP1) This task involves ETHZ, Fraunhofer, CTU, and DFKI and end user organization FDDo.

The purpose of the domain analyses for WP1 (§1.3.5.1) is to construct empirical profiles of sensor behavior. These profiles are based on data we intend to gather by using sensors under different circumstances in training areas (e.g. different types of smoke, heat sources). The sensors to be profiled are those we plan to use for navigation and observation in WP1 and WP2 (§1.3.5.2).

2.3.1.2 Involvement: Domain analyses for spatial dialogue in HRI for USAR (WP3) This task involves DFKI and end user organizations FDDo, VVFF and EiKdo.

The purpose of the domain analyses for WP3 (§1.3.5.3) is to

- have experts collect image and video data of significant terrain features, landmarks and threats typically found in areas of NIST types Yellow, Orange, Red;
- have experts annotate this visual data with descriptions to uniquely identify each feature, landmark or threat found in the data; and,
- use the resulting annotated corpus as a benchmark for what a robot needs to be able to understand and produce when talking about the environment.

For DR 3.1.1 (M12) the domain analysis concerns static phenomena in disaster areas. For DR 3.2.2 (M24) this concerns dynamic phenomena (dynamics in the environment, including threats). For each DR we have a data collection phase (a.1), a data sample identification phase (a.2), followed by a descriptive annotation phase (b.1) and a semantic stand-off annotation phase (b.2). End users are involved in phases (a.1) and (b.1).

For each DR, data collection (a.1) is based on the creation of video material. The video material needs to show sample situations in disaster areas of different types, and be commented by a rescuer "as if" the person is exploring these areas and informing others. The material needs to be sufficiently varied across the area types. Estimated data-size of the material for one training area is about 0.5h video for NIST Yellow, 1h video for NIST Orange, and 1h video for NIST Red. We intend to create material for three different training areas, thus creating 5h-7.5h of video material. Based on the collected data, the walk-through comments are used to identify samples within the video stream in which significant (commented-upon) terrain features, landmarks, or threats are found (a.2). For DR 3.1.1. these samples are images, for DR 3.2.2 these samples are video clips. The resulting sample set should be significantly large to show different instances of similar types of terrain features, landmarks, or threats.

A rough estimation for DR 3.1.1 is that the video material per site yields 100-150 instances. Representative samples will be selected from classes of instances (a.2), to be annotated by experts to construct a seed corpus (b.1). Given 300-450 samples, we plan for about 100 representative samples to be annotated. For each sample image, (b.1) we plan to have 3 to 4 experts provide natural language expressions that identify any terrain features, landmarks, or threats they believe can be identified in the image. These expressions are then transformed into semantic representations, which are used as (stand-off) annotation for the image data (b.2). We intend to investigate whether the seed corpus can be used to semi-automatically annotate the entire corpus of image data. An estimation for DR 3.2.2 is that the video material yields a smaller set of instances, in the order of 50-70 representative video clips. Like the images, video clips will be annotated by descriptions of the observable events.

End users are involved in data collection (a.1), and (b.1), for each DR. Estimated overall efforts for (a.1) is 0.75PM, including access to a training area. The estimated overall end user efforts for (b.1) for DR 3.1.1. are 1PM for a data set of 100 images to be annotated, for DR 3.2.2. we estimate 1PM for a data set of about 50 video clips to be annotated. The distribution of these PMs is given in Table 7. WP efforts estimated for (a.2) is about 1PM per DR, for (b.2) about 2PM (constructing semantic representations) per DR. The resulting annotated data sets provide input to WP2 (acquiring visual models), to WP3 (coverage for comprehending and producing situated dialogue). The data sets will be provided as open source, to provide a specialized benchmark for situated dialogue in HRI.

2.3.1.3 Involvement: Domain analysis for user, task and context modeling (WP4)

This task involves TNO and end user organizations FDDo, and VVFF.

The purpose of this domain analysis for WP4 (§1.3.5.4) is to

- acquire domain knowledge from practitioners in the field and subject matter experts,
- have practitioners and experts validate the resulting knowledge descriptions; and,
- establish a coherent and concise set of user characterizations and needs (operational demands) for human-robot interaction.

In the first phase of knowledge acquisition, subject matter experts provide relevant documentation. Subsequently, open and closed interviews are being conducted with different members of a team (half-a-day session for four team members). We plan to interview members of six teams. In parallel, training material is being collected as far as available, and trainers are being interviewed (one trainer for half-a-day). We plan to interview 6 trainers.

In the second phase, the (concise and coherent) descriptions of the domain knowledge is presented to the practitioners and experts for review. Four review sessions are planned that need preparation from the participants (reading the material and providing their first individual assessment).

In the third phase, specifications on user characterizations and operational demands (such as scenarios), which are relevant for human-robot interaction, are being assessed in a similar way as the domain knowledge descriptions of phase 2.

Results feed into DR 4.2.1 (M16). The estimated end users effort is 3PM.

2.3.1.4 Involvement: HF use case definition and human in the loop testing (WP4)

These tasks involve TNO and end user organizations FDDo, and VVFF.

The purpose of these specifications and evaluation efforts for WP4 (§1.3.5.4) is to

- involve the end-users in the generation of human factor design specifications; and,

- test the theories and models of the prototypes with human-in-the-loop tests.

In Yrs 2, 3 and 4, the end-users are involved in scenario-based assessments. Both explorative (formative) prototype assessments and controlled experiments are conducted to refine and validate the user needs and to establish a sound foundation of the theories and corresponding models.

For the explorative assessments, six practitioners will be involved in two half-day sessions. These participants will also review the reports of the sessions.

In the controlled experiments, about 30 participants will work with alternative versions of the prototype. At the end of the session, the video takes will be replayed and the participants will annotate (assess) their own performance. A subject matter expert will contribute for the preparation of the assessments and by assessing the performance during the test.

Results of these activities will feed into DR 4.3.2 (M34), DR 4.4.3 (M40), and DR 4.4.4 (M48). The estimated effort over 3 years is 3PMs per year, i.e. 9PM in total.

2.3.1.5 Involvement: Domain analysis for skill learning in USAR (WP5) This task involves ROMA and end user organization VVFF.

The purpose is to have experts showing skills for the specified tasks, to be modeled for the robot in T5.1 in WP5 (§1.3.5.5). Showing is to be settled to understand behaviours and ensure clear replication in a coordinate action execution.

End users will provide movies of well-specified sequences of actions. These movies will be made using a skilled rescuer wearing ROMA's Gaze Machine. This machine records video and possibly audio (microphones). The use of the Gaze Machine ensures we cam accurately measure coordination between modalities involved in acting during search. Examples are perceptual switching including gaze direction and relevant focus. Measurements must be taken in round clock operations, for speed, pressure on the terrain, body inclination, sequence of fixations, detailed communication with the operator central and switching time for perceptual modalities. Specification of the hear-see in each instant of the behaviour showing is also provided.

Typical experiments under different perceptual circumstances will be planned and then settled to indicate motion and perception coordination, gestures for alert, focus of attention, region of interest, gaze direction under requests, speed in search, active perceptual modality and sequences of move-stop-run. When indicating regions it is required to use both gaze and coordination headfinger, when switching modality speaking is required.

The individual skills can then be combined into simple maneuvers with other operators, for example in searching, reaching or communicating an alert (for example "what do you see on the left..") for coordination with humans or another robots.

Overall data to be collected:

- Double movies (from outside + inside with the Gaze Machine, annotated with voice) for at least one hour of successful experiments of about 2*90k frames. Movies should provide a detailed analysis of coordinated gesture perception in specific, planned circumstances.

- A dictionary of expected action (situation reaction) and protocol manual for looking at, search, victim search, reporting, loosing communication, etc. The dictionary specifies at least 30 video documented action sequences.

The estimated effort for Task 5.1 (M6) is 3PM for the end user, including interviews, experiments, planning and settings of scenarios. This contributes directly to DR 5.1.1. For Task 5.2. (M11) the estimated effort is 2PM for the end user.

2.3.1.6 Involved: Progression and testing for skill learning (WP5) This task involves ROMA and end user organization VVFF.

The end user is involved in testing mixed initiative, especially for interoperability such as inspection, status reporting and situation awareness. This contributes to Task 5.6 (DR 5.4.6). The estimated effort is 2PM for the end user.

2.3.1.7 Involvement: Specification for rover platform design (WP6) This task involves BLUE and all other NIFTi partners involved in WP6, as well as end user organizations FDDo, VVFF, EiKdo, and RUAG.

The purpose is to involve end user organizations in the specification of the design for the NIFTi rover platform, developed in WP6 (§1.3.5.6). We plan to organize this as a 2-3 day specification meeting, with an additional round of feedback from end user organizations on a draft of the specification document. The estimated overall effort for end users is 1PM, 0.25PM per end user organization. The results are reported in DR6.1.1 (M6).

2.3.1.8 Involvement: Specification and evaluation of integrated systems (WP7) This task involves all the NIFTi partners, including the end user organizations FDDo, VVFF, EiKdo, and RUAG.

The purpose is to involve end user organizations in the specification and evaluation of the NIFTi integrated systems, developed in WP7 (§1.3.5.7). We refer to WP7 for the exact description of the yearly missions, criteria, and targets on which these systems are to be specified and evaluated. The estimated overall effort for end users is 8PM for four years: per year, 1PM per year on specification, 1 PM per year on evaluation.

2.3.1.9 IFR - Institut für Feuerwehr und Rettungstechnologie/Research Institute for Fire Service and Rescue Technology Fire Department of Dortmund (FDDo)

Organization The FDDo is the sixth largest fire department in Germany. It is responsible for fire fighting, rescue and emergency management in the city of Dortmund, which has an area of 281 square kilometre with about 590.00 citizens. 700 professionals (incl. administration and management) run 9 fire stations, about 1000 volunteers organised in 19 fire brigades support the professionals. Inside the city of Dortmund an underground railroad network and two through road tunnel (each four lanes, two tunnels, length 1.700 m rsp. 800 m) an airport and a domestic port are operated. Additionally there are several industrial plants and a technical university established in the City of Dortmund. Therefore the members of the Fire Department Dortmund have great experience and expertise they can contribute to this project. In addition FDDo has experience in scientific projects like the EU-Projects VIRTUALFIRES, SHARE and PRONTO as well as the several national funded Projects.

Due to the named projects and the foreseen additional research needs the city council of Dortmund in 2006 established the IFR - Institut für Feuerwehr und Rettungstechnologie/Research Institute for Fire Service and Rescue Technology as a additional sector of the fire department. Since January 2009 the IFR is established as an local government department responsible for contributing to and coordinating research projects together in cooperation with FDDo (www.ifr.dortmund.de).

Contributions and experience The environment of FDDo is qualified for end user requirement definition and field tests. The close contact to other German and European Fire Departments/organisations is a good basis for system design and validation purposes. Within this project the tasks of the Dortmund Fire Department are focussed on the definition of user requirements and use cases, and the evaluation of project results. For evaluation purposes the IFR/FDDo is able to collect, concentrate and integrate the information of other European Fire Departments and related organisations.

Staff *Dipl.-Ing. Klaus Schäfer*, until end of 2008 Head of Fire Department Dortmund and now full-time Head of IFR - Institut für Feuerwehr und Rettungstechnologie/Research Institute for Fire Service and Rescue Technology. He has been responsible for the Fire Department Dortmund, the Rescue service, the disaster control and civil defence and is still cooperating with several national and international working groups dealing with fire fighting, fire protection and rescue methods, including aircraft and airport fire fighting. From this he has great experience in the areas of risk-analysis, rescue planning and the evacuation planning.

Univ.-Prof. Dr.-Ing. Rainer Koch, Deputy Head of IFR - Institut für Feuerwehr und Rettungstechnologie/Research Institute for Fire Service and Rescue Technology and volunteer fire officer of the Dortmund Fire Department. In his main profession he is head of the research group Computer Application and Integration in Design and Planning (C.I.K.) at University of Paderborn, is professor for optimisation of design- and planning-processes within computer-aided engineering. His applied research is focussed on Virtual/Augmented Reality, database technologies and mobile applications. He has significant experience in Virtual Prototyping, computer based simulation and visualisation of various technical scenarios. Also the C.I.K.-research group has great experience in scientific projects like the EU-Projects VIRTUALFIRES, SHARE and PRONTO as well as in coordinating and contributing to several national funded research projects.

2.3.1.10 Corpo Nazionale Vigili del Fuoco (CNVVF)

Organization The unit NBCR of the Ministry department of Interior is part of the Fire Brigades, of the Public Rescue and Civil defense, and in particular is part of the Central Direction for the Emergency and Technical Rescue, and directly collaborate with the Central Direction of the logistic and instrumental resources. The role of the NBCR area inside the Emergency Central Direction is to control and face the risks concerned with dangerous agents, namely: Nuclear (N), Biological (B), Chemical (C) and Radiological (R). The National Fire Department is in charge of contrasting the risks from chemical agents, since its foundation in the year 1941. From 1961 it also faces the nuclear risks and in the last years it has faced the problems concerned with the biological risks.

Contributions and experience The "Corpo Nazionale dei Vigili del Fuoco" is a National structure of the Interior Ministry (home affairs), that is, it is the Department of Fire Operators, Public Rescue and Civil Defense. The chief of the VVFF is currently Antonio Gambardella. Their contribution to the project is mainly focused on providing the specialized scenarios offering possible areas of tests in Monte Libretti which is a poly-functional area for rescue simulation. The operators will also play a crucial role in providing the expertise, by mean of skilled operators, for learning skills, which is a basic step for the design of execution control and monitoring, flexible planning and mixed initiative planning with situation awareness.

Staff *Dott. Ing. Alfio PINI*, Central Director for the emergency and technical security, expert of chemical plants and interventions in this sectors;

Dott. Ing. Antonio GAMBARDELLA, Central Director for the logistic and instrumental resources, expert of radioprotection.

Dott. Ing. Massimiliano GADDINI, Manager of the area for the N/R risk and responsible of laboratory for the radioactivity control and radioprotection.

Dott. Ing. Marco FREZZA, Manager of the area for the BC risk, national expert of III operative level.

Dott. Ing. Emanuele PIANESE, Vice-responsible of the Area for the N/R risk and expert of radioprotection.

Dott. Ing. Michele MAZZARO, Area Officer for the N/R risk and expert in radioprotection .

Dott. Ing. Paola BLOTTA, Vice-responsible of the area for the BC risk, national expert of III operative level.

2.3.1.11 Einsatzkommando Katastrophenhilfe Bereitschaftsverband (EiKdo)

Organization Throughout the whole year, the headquarters of the **Einsatzkommando Katastrophenhilfe Bereitschaftsverband** EiKdoKataHiBerVb (Military Disaster Relief Rapid Reaction Command, `http://www.he.admin.ch`) provide cover for military disaster relief at home and abroad.

For 22 years now, the command has provided disaster relief within numerous urban search and rescue operations in earthquake regions all over the world. During these years, the lives of 22 people have been saved. In Switzerland, the command's professionally trained and highly experienced soldiers have been deployed in all last years' major disastrous incidents.

The command provides help according to the International Search and Rescue Advisory Group (INSARAG) of UNO/OCHA. Its operational doctrine and its scale of equipment give it the ability to provide world-leading, sophisticated support to the civilian emergency services quickly and professionally without in any way competing with them. When not engaged on disaster relief work, the formation trains, provides support and carries out tasks for the benefit of third parties.

Contributions and experience The expertise of the command in carrying out disaster relief missions will serve as an invaluable source of information to the conceptualization of demonstrator scenarios in NIFTi. Furthermore, the partners and EiKdo plan to closely interact on the development and deployment of rescuing robotics hardware. Here, EiKdo's practical experience in handling state-of-the-art technical devices will be particularly relevant when identifying requirements and construction constraints for new platforms.

The Swiss army possesses three well-maintained training facilities that have exclusively been built for search and rescue training purposes. These three facilities are situated in Genf, in Wangen aA, and in Berhardzell near St. Gallen and can be made available to the partners for experimental purposes. Thus, complete systems or components may directly become involved in field tests, under the experienced supervision and with the help of EiKdo.

Staff 11 military professionals and 1 civilian employee work in the command and are available as competent contacts with expertise in their field. Oberstlt i Gst Christen is the head of the organization. In association with his assistant, Major i Gst Matthias Pfister, he will therefore serve as the main contact person for NIFTi.

The disaster relief standby company represents a combination of soldiers from the rescue and engineer corps, which are exclusively led by military professionals, contracted military personnel and civilian employees. It consists of approx. 150 single term conscripts and approx. 50 contracted military persons. With three schools beginning each year, this results in three companies with a military personnel strength of approx. 200 each.

Usually only those soldiers are assigned to the disaster relief standby company as single term conscripts who have done their 18 week recruit school with engineer or rescue corps. Military cooks and motor vehicle drivers are naturally also assigned to this unit.

2.3.1.12 RUAG Land Systems (RUAG)

Organization RUAG Land Systems is an internationally renowned company for defence and security products and services. Research, development and testing are specialized on the needs of mobile and immobile missions as well as for the appropriate equipment. As industrial partner of the Swiss Army we are integrating communication and reconnaissance systems into mission vehicles and centres. Further, we are maintaining a competence centre for tracked vehicles at the headquarter in Thun. We have a large experience in mechanical and electronic systems engineering, design and manufacturing of mobile systems. Besides the performance upgrade of vehicles in service, we also transform their structure and logistics to new demands. Among recent results are a remote controlled, 15t-vehicle for the demining of rural surfaces, an autonomous Laser spot tracking system and an autonomous controlled car from the ETHZ (Prof. Siegwart) upgraded for the necessary electric power and safe operation equipment. The company consists of about 400 employees. A team of 6 scientists is focusing on applied research topics which form the base of our future technology performance. Supported by the mechanical and electronics development department, engaging 150 engineers (CAD and design) and technicians, marketable research results are usually investigated by means of operational models as well as prototypes. Additionally, our testing department is exposing those systems to environmental compatibility tests including even electromagnetic compatibility. With respect to its clients from defence and security the company is certified according to ISO 9001, ISO 14001, AQAP 2110 and ISO 3834-2.

Role and commitment of RUAG Land Systems to NIFTi The role and commitment of RUAG Land Systems to NIFTi consists of the system integration support, environmental testing and operational testing in the field in urban and rural scenarios. For the latter tests, we base our contribution on a very large experience in ”out-of-the laboratory” testing including preparation of the necessary out-door-equipment. Further we can serve the project with a wide knowledge of many interesting testing ranges in Switzerland.

Partner	Key contributor	Project	%FTE
DFKI	Geert-Jan M. Kruijff (coordinator)	NIFTi	50%
		CogX	30%
		ALIZ-E	20%
	Hendrik Zender (leader WP3)	NIFTi	80%
		CogX	20%
	Remaining FTE (approx.)		1.5
TNO	Mark Neerincx	NIFTi	12%
		ALIZ-E	7%
	Rosemarijn Looije	NIFTi	24%
		ALIZ-E	27%
	Chris Jansen	NIFTi	10%
	Jurriaan van Diggelen	NIFTi	21%
	Remaining FTE (approx.)		0.55
Fraunhofer	Rainer Worst	NIFTi	100%
	Hartmut Surmann	NIFTi	50%
	Remaining FTE (approx.)		1.0
BLUE	Nicola Tomatis	NIFTi	5%
		BACS	5%
		BRICS	5%
		EUROPA	5%
		robotshome	5%
	Pierre Lamon	NIFTi	10%
		BACS	5%
		EUROPA	10%
		robotshome	5%
	Gregoire Terrien	NIFTi	10%
		EUROPA	10%
		robotshome	10%
	Remaining FTE		0.375
ETHZ	Roland Siegwart	NIFTi	5%
		Others	30%
	Cedric Pradalier	NIFTi	15%
		Robots@Home	25%
		sFly	15%
	Francis Colas	NIFTi	80%
	Remaining FTE (approx.)		1.5
CTU	Václav Hlaváč	NIFTi	15%
		HUMAVIPS	15%
	Tomáš Pajdla	NIFTi	10%
		HUMAVIPS	10%
		ProVisG	40%
	Tomáš Svoboda	NIFTi	30%
	Karel Zimmermann	NIFTi	50%
	Remaining FTE (approx.)		0.30
ROMA	Fiora Pirri	NIFTi	35%
		ViewFinder	2%
		COSMO-SkyMed	15%
		Others	5%
	Marco Schaerf	NIFTi	15%
		COSMO-SkyMed	10%
	Remaining FTE (approx.)		1.00

Table 3: Involvement of key project contributors (in %FTE) on NIFTi and ongoing projects

Approved by EC on 2009-12-09

Partner	Expertise	Contributions to research	Obj.
DFKI	Human-Robot Interaction	Spoken dialogue for HRI	3
		User, context-driven variation of dialogue production	2,3
		User, context-driven robust spoken dialogue comprehension	2,3
	Intelligent user interfaces	User, context-adaptive GUIs for multi-modal HRI	2,3
	Cognitive architectures	Cognitive architecture design, integration	–all–
		Symbol grounding of conceptual structures	1,3
	Robotics, mapping	Domain inference, functional projection in conceptual mapping	1
	User psychology	Measures for linguistic processing load	2,3
TNO	Cognitive engineering	Domain analysis, quantitative models for user modeling	2
	Human-Robot Interaction	Domain analysis, interaction models, quantitative evaluation of system design	2,3
	Human-centered design	Domain analysis, quantitative evaluation	–all–
	User psychology	Modeling of cognitive task load, situation awareness and trust	2
	Intelligent user interfaces	Working agreements for shared control, user-adaptive GUIs	2,3
Fraun-hofer	Robotics, design & control	Design, integration, architecture of complex hard- & software	1,4
		6DoF autonomous robots control, robot rescue; UAV	1,4
	Robotics, mapping	2D/3D mapping, 6D-SLAM; benchmarks	1
	Computer vision	Sensor fusion, visual attention	1
BLUE	Robotics, design & control	Development of a new adaptive locomotion concept	4
		Hard- & software architecture for robot control in rough terrain	4
ETHZ	Robotics, mapping	Spatio-temporal world representation layers for dynamic grounding of higher-level concepts	1
		Fusion of visual mapping modalities with spatio-temporal representations into hybrid maps	1
	Robotics, design & control	Trajectory control, obstacle avoidance based on spatial models	4
CTU	Computer vision	Visual spatial and temporal information for mapping	1
		Statistical and structural learning on image/video data	1
	Cognitive architectures	Symbol grounding using image understanding techniques, generalizing percepts to concepts	1
ROMA	Planning and execution	Task-load driven planning for joint exploration	2,4
		Adaptive mechanisms for planning and execution monitoring, with varying shared control	4
	User psychology	Multi-modal perceptual attention, dynamic saliency map	2,4
	Computer vision	Multi-view data association, silhouette recognition	1

Table 4: Partner expertise and contributions to the project objectives

Approved by EC on 2009-12-09

Objective	Required competences	Partner competences
Obj.#1	Outdoor spatial modeling	ETHZ, Fraunhofer,
	Spatio-temporal modeling	ETHZ
	Vision and sensory processing	CTU, Fraunhofer, ROMA
	Functional environment models	CTU, DFKI, ETHZ
Obj.#2	Cognitive modeling	TNO, DFKI
	Human factors and evaluation	TNO
Obj.#3	Human-robot interaction	DFKI, TNO
	Multi-modal dialogue systems	DFKI, TNO
	Spoken dialogue processing	DFKI
	Time-critical language processing	DFKI
Obj.#4	Field robotics	ETHZ, BLUE, Fraunhofer
	Robot control	ETHZ, BLUE, Fraunhofer
	UAV	ETHZ, Fraunhofer
	Flexible planning	ROMA, DFKI
	Skill acquisition	ROMA, DFKI
	Attentional models	ROMA, DFKI, TNO
Overall	Cognitive architectures	DFKI, ROMA
	Symbol grounding	DFKI, CTU, ROMA

Table 5: Competence matrix, project requirements vs. partner competences

Competence	DFKI	TNO	Fraunhofer	BLUE	ETHZ	CTU	ROMA
Field robots, varying morphology			X	X	X		
Field robots, robot control			X	X	X		
UAV			X		X		
Cognitive architectures	X						X
Symbol grounding	X					X	
Cognitive modeling	X	X					
Human factors & evaluation		X					
Outdoor spatial modeling			X		X		
Spatio-temporal modeling					X		
Functional environment models	X				X	X	
Vision & sensory processing			X			X	X
Human-robot interaction	X	X					
Multi-modal dialogue systems	X	X					
Spoken dialogue processing	X						
Time-critical language processing	X						
Flexible planning	X						X
Skill acquisition	X						X
Attentional models	X	X					X

Table 6: Competence matrix, organized by partners

End user	WP	DR	Month	Activity	PM	Tr. area access	Description
	WP1	DR1.1.1	M12	Data collection	**0.25**	yes	Sensor-profiling in training areas
FDDo					0.25		
	WP3	DR3.1.1	M12	Data collection	**0.75**	yes	Image/video of static landmarks, features, threats
EiKdo					0.25		
FDDo					0.25		
VVFF					0.25		
	WP3	DR3.1.1	M12	Knowledge transfer	**0.75**	no	Expert annotation of subset of image/video data
FDDo					0.25		
VVFF					0.50		
	WP3	DR3.2.2	M24	Data collection	**0.50**	yes	Image/video of dynamic landmarks, features, threats
EiKdo					0.25		
VVFF					0.25		
	WP3	DR3.2.2	M24	Knowledge transfer	**0.75**	no	Expert annotation of subset of image/video data
FDDo					0.25		
VVFF					0.50		
	WP4	DR4.2.1	M16	Knowledge transfer	**2.50**	no	Evaluation of user model (task load, attention)
FDDo					1.00		
VVFF					1.50		
	WP4	DR4.3.2	M34	Evaluation	**3.0**	yes	Evaluation of working agreements, HRI adaptation
FDDo					1.50		
VVFF					1.50		
	WP4	DR4.4.3	M40	Evaluation	**3.0**	yes	Evaluation of working agreements, HRI adaptation
FDDo					1.50		
VVFF					1.50		
	WP4	DR4.4.4	M48	Evaluation	**3.0**	yes	Evaluation of adaptive HRI
FDDo					1.50		
VVFF					1.50		
	WP5	DR5.1.1	M6	Knowledge transfer	**3.0**	yes	Requirements for robot skill primitives: planning
VVFF					3.0		
	WP5	T5.2	M11	Knowledge transfer	**2.0**	yes	Requirements for robot skill primitives: learning
VVFF					2.0		
	WP5	DR5.1.1	M10	Evaluation	**2.0**	yes	Evaluation of mixed-initiative cooperation
VVFF					2.0		
	WP6	DR6.1.1	M6	Specification	**1**	yes	Requirements for platform
RUAG					0.25		
EiKdo					0.25		
FDDo					0.25		
VVFF					0.25		

Table 7: End user involvement in WPs 3–6.

End user	WP	DR	Month	Activity	PM	Tr. area access	Description
	WP7	DR7.1.3	M12	Specification	**1**	yes	Requirements for human-guided exploration
EiKdo					0.25		
RUAG					0.25		
FDDo					0.25		
VVFF					0.25		
	WP7	DR7.1.3	M12	Evaluation	**2**	yes	Evaluation of system for human-guided exploration
FDDo					0.50		
VVFF					0.50		
	WP7	DR7.2.4	M24	Specification	**1**	yes	Requirements for human-assisted exploration
EiKdo					0.25		
RUAG					0.25		
FDDo					0.25		
VVFF					0.25		
	WP7	DR7.2.4	M24	Evaluation	**2**	yes	Evaluation of system for human-assisted exploration
FDDo					0.50		
VVFF					0.50		
	WP7	DR7.3.5	M36	Specification	**1**	yes	Requirements for in-field joint exploration
EiKdo					0.25		
RUAG					0.25		
FDDo					0.25		
VVFF					0.25		
	WP7	DR7.3.5	M36	Evaluation	**2**	yes	Evaluation of system for in-field joint exploration
FDDo					0.50		
VVFF					0.50		
	WP7	DR7.4.6	M48	Specification	**1**	yes	Requirements for sharing situation awareness
EiKdo					0.25		
RUAG					0.25		
FDDo					0.25		
VVFF					0.25		
	WP7	DR7.4.6	M48	Evaluation	**2**	yes	Evaluation of system for sharing situation awareness
FDDo					0.50		
VVFF					0.50		

Table 8: End user involvement in WP 7.

2.4 Resources to be committed

The total requested EC contribution to the project budget is about 6.6M euros. About 5.7M EUR covers RTD personnel costs, for close to 55 person years of effort (75% EU contribution) for RTD activities. Although the total person months tend towards the lower end of the range of total person effort for large scale integrating projects, we believe we can achieve a more effective use of these efforts by virtue of the smaller project size. This allows for a tighter participant integration. Non-personnel costs cover subcontracting, the purchase and development of specialized equipment, the dissemination plan and community building, and management. These costs are justified in more detail below.

Table 9 lists for each partner the direct costs per person month (PM) for different activities (RTD, MGT, OTH), per partner. OTH costs are not listed for FDDo, VVFF as they have no such involvement in NIFTi.

Partner	RTD (75%)	MGT (100%)	OTH (100%)
DFKI	7200	7200	7200
TNO	6714	6714	6714
Fraunhofer	6096	7500	5870
BLUE	8900	8900	8900
ETHZ	6000	6000	5850
CTU	5600	5600	5600
ROMA	6700	6700	6700
FDDo	6200	5400	–
VVFF	6500	6500	–

Table 9: Direct costs per PM for RTD, MGT, and OTH (in EUR)

The 8900 EUR/PM are calculated by taking BLUE's average personnel costs.

2.4.1 Receipts

The partner DFKI plans to use resources made available by third parties (i.e. the states of Rhineland-Palatinate and the state Saarland). The resources include salaries of professors and researchers paid by the governments, as well as equipment, infrastructure and services paid by the governments. The total amount of such receipts will be 159.192 Euro and charged as percentage of own personal costs.

2.4.2 Resources for subcontracting

The budget plans a total of 30k euros for subcontracting a rescue organization (EiKdo, 15k EUR), and a company with experience in system integration and -evaluation in the area of rescue robotics (RUAG; 15k euros). The organizations will have limited involvement in activities concerning data collection and specification in WPs 3, 6, and 7. Subcontracting costs will primarily cover travel costs and training area use. Personnel costs will be covered by the participating organizations themselves. The precise involvement is described in §2.3.1, and is summarized in Table 7 (p.110) and Table 8 (p.111).

2.4.3 Resources for specialized equipment

The budget plans a total of 638K euros for specialized equipment. The consortium brings existing hardware with a significantly higher value to the project. This includes several high-end mobile outdoor robot platforms such as a TeleMAX 500, and multiple VolksBot, Pioneer P3-AT, BlueBotics Shrimp and Crab platforms, and a wide variety of sensor platforms:

DFKI contributes a well-equipped robotics lab. The lab provides a mixed-platform environment (Mac, Windows, Linux) and includes several Pioneer robots with a wide variety of sensors (>150K EUR). The robots will be made available to NIFTi for cross-platform experimentation (e.g. P3-AT).

TNO TNO has a robotic lab with multiple robots and different interface options (e.g. telepresence, haptic). The robotic lab including equipment has a magnitude of > 500K EUR.

Fraunhofer IAIS has a robotic lab with mechanical and electrical machines and equipment. For robotic experiments in the first two years the Volksbot robots will be used. The robotic lab including equipment has a magnitude of 1 Million Euros.

ETHZ contributes a well-equiped robotics lab including several hardware plateforms, workstations, and state-of-the-art sensors available for data sources and experimentation. Among those are a Velodoyne sensor as well as other rotation laser scanners (60k EUR), a hybrid car equipped with a localisation system (70k EUR), and several space rovers (70k EUR).

CTU provides to the project the background of the well equipped computer vision lab including the optical bench, free optics, cameras (including far infra red one by FLIR in the 8-10 μm wavelength range), lasers, etc. We have also three industrial robots available in a specialized lab. CTU runs grid of computers for computational power demanding tasks.

ROMA contributes a well-equipped robotics lab including a Pioneer P3-DX, a P3-AT and a Bluebotics Shrimp robotic platforms; also different kinds of sensors, workstations and acquisition devices are contributed: two Hokuyo LRFs, audio acquisition facilities, a Gaze Machine for 3D gaze tracking, a CUDA based high performance computing workstation, two pan-tilt units, two inertial measurement units and a number of cameras.

FDDo contributes access to, and the use of, a large training area in Dortmund.

VVFF contributes access to, and the use of, a large training area in Monte Libretti.

Given NIFTi's goal to develop highly autonomous, interactive cognitive robots, it is crucial for the project to go beyond the state-of-the-art, and extend and develop hardware solutions.

NIFTi develops a prototype platform that enables the investigation of how to answer the need for better locomotion systems, with extended variable morphology (WP6). The platform must demonstrate high climbing performance for better scene coverage. As the platforms available on the market do not fit our needs, the first prototype we propose to build is a new rover structure. This new design is produced at several examplars and distributed to partners of the project. The design of a new rover has two main advantages: (1) the integration is much faster/easier because all partners work on the same hardware, and (2) a new design allows to account for the partners and end-users requirements (payload, size, sensing capabilities etc.). NIFTi plans the cost of each prototype to be maximally 50k euros. The robots are equipped with different sensors (such as cameras, microphones, gas sensors, etc.) for inspecting the environment, and a robot arm. The market offers several off-the-shelf arms (20k euros) that can be used for the project and there is no need to develop a specific design. Table 10 details the estimated costs per rover platform. Table 11 specifies the platform costs per partner.

Furthermore, NIFTi plans to acquire Unmanned Aerial Vehicles (UAVs), to be used to improve the overall situational awareness. The challenge here is to integrate consistently the data coming from the UAVs with data generated on the ground. Because NIFTi does not focus on the UAV development and control, off-the-shelf units will be bought. NIFTi plans a cost for 4 UAVs, with a unit price of 10k euros, including the embedded controller and the sensing equipment. The overall budget for UAVs is 40k euros.

Platform component	Est. cost (in EUR)
Rover	50.000
Sensorsuite (total)	25.500
– *Pan-Tilt Unit for* $3D^+$ *mapping (WP1)*	4000
– *Laser range finder (WP1)*	4000
– *Pan-Tilt Unit for stereo camera (WP2)*	800
– *Stereo camera (WP2)*	2000
– *Omni-directional camera (WP2)*	3000
– *On-board microphone array (WPs 2,3)*	1700
– *Infrared camera(WP2)*	10000
Robot arm	13.500
Total	**89.000**

Table 10: Estimated costs per rover platform, over the entire development (in EUR)

Platform	DFKI	TNO	Fraunh.	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Rover	89.000	89.000	89.000	89.000	74.000	79.000	89.000	0	0
UAV	10.000	10.000	10.000				10.000	0	0
Total	99.000	99.000	99.000	89.000	74.000	79.000	99.000	0	0

Table 11: Estimated platform costs per partner (in EUR)

2.4.4 Resources for partner exchange

It is important for the NIFTi project to build up and maintain a tight integration among the consortium participants. Therefore, the project will support research integration through a partner exchange program. Financially, the budget plans a total of 63k euros for this, cf. Table 12. This yields about 1.1k euros which each partner can spend every year on partner exchange, per person year of RTD effort.

Exchange	DFKI	TNO	Fraunh.	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Total budget	11.250	9.000	9.000	4.500	11.250	9.000	9.000	0	0

Table 12: Partner exchange budgets per partner (in EUR)

2.4.5 Resources for dissemination and community building

NIFTi intends to maintain a high profile internationally, by striving to publish in high impact journals and at major conferences such as ICRA, IROS (robotics), AAAI, IJCAI (artificial intelligence), HRI, and RO-MAN (human-robot interaction). This ensures maximum visibility and potential take-up of project results while keeping the overall costs to a reasonable sum, (taking into account the fact that many of these conferences are typically in North-America or in the Far East). The project budgets a total of 130k euros for dissemination through academic conferences. Table 13 details how this is distributed across partners. Dividing by person years provides an average of about 2.5k euros per person per year (about 2 international conferences) for each partner.

Travel	DFKI	TNO	Fraunh.	BLUE	ETHZ	CTU	ROMA	FDDo	VVFF
Total budget	25.000	20.000	20.800	5.000	20.000	20.000	20.000	14.000	13.200

Table 13: RTD travel budgets per partner (in EUR)

Furthermore, NIFTi organizes a yearly summer school (DFKI, TNO, CTU and ETHZ). The summer schools are used to strengthen the community within the consortium, and provide an outreach to other related communities. Each summer school lasts approximately one week, to be attended by the researchers on the project and up to 20 researchers from other communities. The goal of each summer school is two-fold. First, leading scientists give tutorials on topics relevant to scientific challenges addressed in the upcoming project year. Second, researchers gain hands-on experience with project components, and integration methodology, by working in teams on assignments set beforehand. Teams are mixed to ensure cooperation with researchers from other project partners. The summer schools have subsidized rates for outsiders to encourage participation; project participants contribute to the costs of the summer school through a moderate registration fee. The projected costs for each summer school, assuming 40 participants, is 15k euros. The majority of the costs are for room and board participants and tutorial lecturers, travel expenses of tutorial lecturers, and for lecture materials. Any hardware necessary for the hands-on exercises is provided by the project partners, to ensure that experience is built up with the hardware actually being used in the project.

2.4.6 Resources for management

The budget plans 2PM per partner for management (except for FDDo with 0.8PM and VVFF with 0.5PM), and 10PM for project coordination (23.3 PMs for management in total). Management costs include per-partner audit costs and 6k euros for travel to management meetings. Furthermore, the DFKI budget includes 15k euros for the SAB. This enables the project to meet several times with the members of the SAB, and cover their travel and accommodation costs.

2.5 Other issues

- As a measure of reinforced monitoring of BlueBotics SA (beneficiary No. 4, BLUE), the Coordinator will receive and manage the pre-financing for BlueBotics. Pre-financing will be paid to Bluebotics in equal instalments every three months upon acceptance by the Project Coordinator and the European Commission of a progress report and work delivered.

- The parts and materials necessary for the robotic platforms will be ordered by Bluebotics which will invoice them with no extras, overheads or profit to the other beneficiaries. These beneficiaries will report the costs to the European Commission. However, the upper funding limits provided for in paragraph 1 of Article II.16 of the Grant Agreement shall be adjusted in such a way as to limit the beneficiaries' reimbursement to 100% of the cost of these parts and materials.

Approved by EC on 2009-12-09

3 Potential impact

3.1 Strategic impact

3.1.1 Strategic scientific impact

Integrated scientific foundations NIFTi puts the human factor into cognitive architectures. It develops the methodology for designing the functionality and infrastructure in a cognitive architecture, to balance operational and cooperation demands across the architecture to minimize a human's cognitive task load and optimize joint work flow. Integration of the scientific foundations is achieved in several ways. The NIFTi consortium has a strong commitment to integration. NIFTi adopts an integrated development cycle (§1.3), and involves all partners in integration and evaluation (WP7). Component prototypes developed by individual WPs (WPs 1–6) are included in the deliverables in the integration WP (WP7) to further strengthen integration. Finally, the NIFTi approach is evaluated at component- and system-level.

Quality of service NIFTi addresses the issue of making human-robot cooperation more natural – and with that, squarely places itself within the aim of improving quality of service. NIFTi improves the services rendered by a robot in cooperation with a human through improved intelligent context- and user-adaptivity.

Innovation capacity NIFTi develops the principles for integrating human factor considerations into cognitive architectures. Despite long research on human factors in human-computer interaction, these insights have not yet been applied to actual cognitive architecture design, nor have they been extended to take into account that both human and robot are mobile, active agents influencing the environment. NIFTi addresses these issues, in an integrated fashion. The resulting innovation is that whenever an application requires a robot to assist humans in performing real life tasks, NIFTi's approaches provide a firm basis for making such assistance natural. After project ending, immediate transfer to closely related domains such as other forms of rescue, safety, and security is envisionable. Furthermore, transfer to service robotics in domestic or public environments is also possible. The NIFTi approach is extendable to include additional factors like gender, age. Gender and age are further aspects of cognitive user models. These extensions can build on the NIFTi architecture as it provides a fundamental approach to modeling the information flow between such cognitive user models and the behavioral processes with which they interact. Notably, factors like gender and age can have an impact on the design and required adaptivity of a dialogue system, for example on acoustic analysis and speech recognition [171, 221, 169] and adaptation of communication [103]. Addressing such factors in the user models, with their relations to cognitive capacities and usage preferences, can improve the foundation for personalization of the user interface [180]. Brief computer-based cognitive tests can be used to refine such models (e.g., age- and gender-related mental rotation tasks; cf., [34]).

Impact of carrying out NIFTi at a European level. NIFTi is interdisciplinary by nature, and it combines basic research with applied technical development. The strategic impact of running NIFTi on a European rather than national scale lies in dissemination, both on academic and on application grounds. In all disciplines touched by NIFTi, i.e. cognitive science, artificial intelligence, user modeling, human-robot interaction, robotics, and USAR, the national academic scenes in each and every EU country are too small to warrant a satisfying forum for dissemination of a novel scientific to making human-robot cooperation more natural. Running NIFTi on a European level helps to address this issue. Furthermore, even at a European level natural human-robot cooperation is not yet addressed directly in previous or existing projects funded by the Cognitive Systems unit. Table 14 places NIFTi in the EU cognitive systems research context, (based on project information available from their respective websites).

Operating at a European level rather than national scale is necessary to maximize chances to exploit the NIFTi results in industry or SME applications. The market potential for natural human-robot cooperation lies not only in USAR, safety or security applications but also in ap-

What?	Who is doing it?	How NIFTi goes beyond
Cooperation	**CO3 AUVS**: "coordination & cooperative control for multiple AUVs" **CHRIS**: "cooperation on manipulating tasks in locally shared environments"	**Putting the human factor into cognitive architectures**: NIFTi integrates cognitive user models into the entire cognitive infrastructure: with how an environment is understood, and how action and interaction are executed and adapted, for natural cooperation in dynamic environments beyond shared locality.
Situation awareness	**CO3 AUVS**: "cooperative situation awareness" **EUROPA**: "probabilistic scene interpretation for changing environments, including detecting and tracking moving objects" **SFLY**: "swarms of μUAVs for USAR-like tasks"	**Cognitive architectures with a genuine notion of situation awareness**: NIFTi couples cognitive user models with plans and a functional understanding of the environment, for a cognitive architecture to interpret and anticipate changes in cognitive user behavior (task load, attention) relative to how the exploration of an environment is proceeding. Comprehension of the environment goes beyond understanding structure: NIFTi projects structure to how it affects cooperation.
Situated communication	**CogX, CoSy, Cogniron**: situated spoken dialogue about tasks in indoor environments, with reference to / description of highly structured environments	**Cognitive architectures which adapt, align communication to a user**: NIFTi considers multi-modal communication (spoken dialogue, GUI) about semi-structured environments, and planning actions in such (partly) unknown environments. Communication does not follow a fixed style, but can adapt in scope, content and form to align the presentation of information to a user's cognitive task load and attention.
Flexible planning & attention	**EYESHOTS**: "object recognition, dynamic shifts of attention, 3D space perception including eye and arm movements, and including action selection in unstructured environments" **IM-CLEVER**: "cumulatively learn new efficient skills through autonomous development based on intrinsic motivations, and reuse such skills [..]"	**Cognitive architectures which learn how to adapt action to the environment (autonomy, morphology) and to a user (cooperation)**: NIFTi drives the acquisition of new skills and skill-adaptations by balancing operational requirements and cooperation requirements. The optimality of operational strategies is considered in the context of joint exploration, what to pay attention to balances information needs of the architecture and of the human.
Morphology adaptation	**LOCOMORPH**: "generate novel and optimal robotic designs which exploit the physical dynamics emerging from the interaction among the physical morphology, control, and environment"	**Cognitive architectures which couple functional maps, plans to morphological adaptation**: NIFTi couples robot control, functional environment models, and exploration plans so the cognitive architecture can anticipate, and effect, changes in its morphology to (a) optimally traverse terrain (locomotion), and (b) perceive terrain features.

Table 14: Placing NIFTi in the EU Cognitive Systems research context

plications in service robotics. The current market volume in service robotics in every single EU country is not of the size that it is safe to assume that a good scientific idea can be turned into applications on purely domestic grounds. Going on a European level is the only way to ensure within a short time frame a critical mass of potential end-users of NIFTi technology.

NIFTi will be firmly embedded into the scientific communities nationally, on a European as well on an international scale. Several partners are active in the euCognition network, and the Europe Robotics Platform (EUROP). NIFTi will strive to ensure that for each partner at least one PI participates in such networks, and where possible that there is a general participation of the project "as entity." All partners are involved in projects funded by the EC under the FP7 Cognitive Systems and Advanced Robotics programs. Exchange between NIFTi and these communities in general will be ensured. Where possible, there will be an active exchange or reuse of results (notably, building forth on software developed in previous projects, e.g. the architecture toolkits from CoSy, CogX).

Interaction with ALIZ-E. Both DFKI and TNO are also involved in ALIZ-E. The goal of ALIZ-E is to develop adaptive systems capable of interaction with a human user (an 8 year-old child) over a longer period of time (circa 5-10 days). There is a certain similarity between ALIZ-E and NIFTi. Methodologically, both projects are concerned with context-sensitive dialogue processing. Both projects require such processing to be robust, adaptive to a user and the communicative context, deal with references to aspect of the environment, and ground content in an understanding of that environment (DFKI). Similarly, these projects share a focus on situated cognitive engineering. Both projects deal with the question, how cognitive user models can be used to enhance the user experience with a robot (TNO).

At the same time, there are important differences between what these projects try to achieve. These differences are ultimately reflected in the tasks to be carried out, (among others by TNO and DFKI). These differences arise first of all from the fact that the projects deal with different domains: expert interaction in urban search and rescue (NIFTi), and child interaction in a hospital ward (ALIZ-E). In these settings a robot plays crucially different roles relative to the user. This is reflected in different kinds of dialogue and spoken language, different types of cognition, and different types of adaptivity – and how we evaluate these systems. NIFTi systems are evaluated on performance in relative "chaos." With the ALIZ-E systems we look at how (emotional) well-being during a hospital stay can be improved. Table 15 summarizes the main differences from the viewpoints of dialogue, and cognitive user models.

When it comes to the actual tasks to be carried out, ALIZ-E and NIFTi can share models of information flow in an architecture. Both need to solve how cognitive user models and robot behavior and -communication interact. This is an obvious point where symbiosis can be achieved. At the same time, the differences outlined above indicate that different types of information will flow, arising from different kinds of models needed to address the project-specific settings. The interesting observation here is that these differences are compatible and often complementary, when viewed from a methodological standpoint. A NIFTi robot is clearly an assistant, task-driven. An ALIZ-E robot varies more in its roles, being driven by the aim to establish a social-affective relationship with the user. If we intend to develop robots that can truly work *with* a human, as more-or-less equal partners in a human environment, then both aspects, both roles need to be integrated into a single complex system. This provides perhaps the strongest point of symbiosis between NIFTi and ALIZ-E. Because they share a methodological basis, they can ultimately bring together aspects of cognition into a single architecture that "no project on its own" would be able to achieve to the depths envisioned by NIFTi and ALIZ-E.

3.1.2 Strategic technical impact

Requirements and domain analyses for, and evaluation of, the NIFTi scientific approach for natural cooperation are performed in close cooperation with potential end-users of NIFTi technology.

NIFTi	**aspect**	**ALIZ-E**
assistent in joint exploration	**role**	pet therapy-roles: companion, educator, coach
dynamic outdoor environments, semi-structured/unstructured	**environment**	highly structured hospital environments, regimented daily routines
collaborative dialogue, about tasks and observations in unknown environments	**dialogue**	verbal/non-verbal communication about events set in daily routines
controlled language, high-precision context-sensitive understanding, english	**language**	semi-free dialogue with a child, "any-depth" understanding, italian
task-oriented adaptation to minimize cognitive task load, optimize work flow	**adaptivity**	affective adaptation, both verbally and non-verbally
semi-srtuctured spatiotemporal aspects of unknown environments, plans	**reference, grounding**	reference to events and people, affective states, structured routines

NIFTi	**aspect**	**ALIZ-E**
professionals	**end-user**	children
cognitive task load, situation awareness	**cognition**	affection, bonding
adaptive automation	**behavioral adaptation**	social behavior
assistant in joint exploration	**role**	companion, educator, coach
performance in "chaos"	**evaluation**	well-being in hospital

Table 15: Differences between NIFTi and ALIZ-E from the viewpoint of dialogue processing (top) and situated cognitive user models (bottom)

This makes a transfer from scientific results to practical applications possible. To exploit this possibility, the proposal includes several business plans for post-project commercialization. There is no doubt that soon more and more applications will be ready for a broad up-take of robotic technology. Currently the US military has the clear lead in these markets with a few but relevant companies. Many of these still are rather small enterprises founded at universities as spin-off companies. The consortium has experience with turning research results into spin-off companies (e.g. DFKI, BlueBotics and Fraunhofer) and will closely monitor exploitation potentials from the NIFTi project: its approach to making human-robot cooperation natural, and developing novel hardware for outdoor applications. BlueBotics clearly spots its interest in the co-development and commercialization of a new rover able to catalyze the new developments of NIFTi. The market potential for civilian applications will be evaluated during the project, but the preliminary market feedback is already very promising. Furthermore, the NIFTi cognitive architecture will provide a new robust reference architecture. It will be made open source at project completion.

To strengthen the links between industry and academia, NIFTi plans to organize yearly Rescue Days, to which industry and academia will be invited. NIFTi fosters an open exchange between the consortium, industry, and academia, and will use the Project Portal (WP8) to set up a networked community to make this possible.

3.2 Plan for the use and dissemination of foreground

The natural way of disseminating basic research results will be pursued, namely, publications in relevant top conferences and journals, visits to research centers, and giving invited lectures. Results will also be presented in the FP7 cognitive systems community where relevant, such as meetings of networks financed by the Cognitive Systems Unit. Since NIFTi is an interdisciplinary project mainly four communities will be addressed, namely artificial intelligence, robotics, human-robot interaction, and the USAR communities. Furthermore, we expect results that are relevant for the fields of computational linguistics, computer vision, learning, and cognitive science. The

NIFTi partners will make their best efforts to ensure that electronic copies of publications become freely and electronically available through the NIFTi portal. This will be done immediately if the scientific publication is published "open access," i.e. if an electronic version is also available free of charge via the publisher, or within 6 months of publication. The related projects and, in particular, research networks addressed below will also be used to disseminate results as well as import as early as possible relevant results from related projects. Finally, DFKI will set up and maintain a web portal that will make all public material (publications, deliverables as far as they are public) easily available. Software, wherever suitable and appropriate, will be made Open Source, (but note the consortium agreement).

Beyond these regular distribution channels, NIFTi will make the following targeted efforts to ensure visibility and wide availability of its results.

Public workshops. From within the consortium, special workshops about interactive situation aware cognitive robots will be organized in the frame program of relevant international conferences, aiming at about one such workshop per project year.

Trade fairs. The project will strive to present project results at international professional trade fairs focusing on USAR-relevant technology (e.g. Hannover Messe) from the second project year onwards, once a year.

Publication of benchmarks. A set of suitable demo problems in the USAR area will be described and made public over the website in such a way that others can reproduce the problems easily and test their approaches easily in comparison to interactive situation aware cognitive robots. The final demonstrators will have to solve these problems. The demo problems will be coordinated with the current activities of the NIST[8] e. g. in the RoboCup Rescue competition. Fraunhofer has already joined the technical committee. Some partners (ROMA, Fraunhofer) already successfully participated in the competition in the past. The tasks during the competition will be more and more realistic.[9] Events together with real rescuers were accomplished during RoboCupRescue 2006 in Bremen which will be extended in the future.

Rescue day. To address industry, and specifically rescuers who are normally not present at conferences or workshops, NIFTi will organize each year a Rescue day starting in the second year of the project. Its purpose will be to present results and technology that are in a state to be exploited towards products or prototypes. Building up and preserving the relevant contacts will go on as a task during the whole project run time. Again, the related researcher networks will be exploited here, for mutual benefit.

Open source. All partners in the consortium are committed to making the knowledge generated during the course of the project as widely and as freely available as possible. Research software developed within the project will be made available as open-source, under a suitable open-source license. For handling IPR issues see the consortium agreement.

Innovation-related activities. Beyond generating scientific knowledge, it is foreseen that NIFTi will also generate technology along interactive situation aware cognitive robots that will be suitable for marketable innovative applications, e.g., rescue robots. To this end, the partners will see to apply for a small number of key patents to give interested European industry and rescue organizations a window of opportunity to develop products under the protection of these patents in succession of the NIFTi project. Furthermore, there are two exploitation plans (§3.2.1). For handling IPR issues see the consortium agreement.

3.2.1 Exploitation plans

The NIFTi project has formulated plans for possible commercial exploitation of the rover platform, and technology for multi-modal human-robot interaction. For confidentiality reasons, the business plan for these exploitation plans will be formulated outside the context of the project – no budget

[8] `http://www.nist.gov/`

[9] `http://www.isd.mel.nist.gov/projects/USAR/Reality_Arena/Reality Arena - NIKE Missile Silo Overview (2005-v1).pdf`

will be reported for these efforts. Market analyses performed for these exploitation plans will feed back into the project, and will serve as input to establishing the community around NIFTi, for example through invitation to the Rescue Days.

3.2.1.1 BLUE: Exploitation of adaptive rover platform Today, BlueBotics is mainly active in mobile robotics for:

- Industrial logistics – Navigation for Automated Guided Vehicles and compact vehicles for light load logistics.
- Service robotics – Prototyping to bring innovative service robotics applications to the market.
- Space robotics – Development of rover and rover components for space robotics.
- Mechatronics – Design, prototyping, and production for innovative mechatronics products.

The company proposes several products including the robot called Shrimp III, a small mobile platform for rough terrain. By commercializing this platform, which is mainly used for research, BlueBotics has learned that there is a clear market pull for a larger rover based on similar characteristics, which are:

- High performance – The climbing ability of the Shrimp III is very high with respect to its modest size.
- Innovation – The proposed solution combines several approaches (parallel bogies, spring frontal fork, etc.) resulting in a innovative solution.
- Simplicity – The passive structure automatically adapts to the terrain without any passive control.

BlueBotics plans to answer to this market pull with the development of the NIFTi rover. As explained within the proposal, contrasting to the space robotics solutions, all the products on the market covering outdoor robotics are based on heavy, active, caterpillar systems. The NIFTi rover will combine the advantages coming from space robotics solutions to the classical approaches of outdoor rovers. The goal of BlueBotics is therefore to bring the NIFTi platform to the market as product. The company will first propose it as outdoor rover for research. Then, the goal is to bring the NIFTi platform and/or the NIFTi components to real world service robotics applications for rescue applications, firefighting, and police applications. BlueBotics will first focus on the commercialization of the intellectual property of BlueBotics only (pre-existing and developed within the project). Then, on potential joint ventures passed on the intellectual property generated by the partners will be investigated (user interface, etc. as resulting from NIFTi). The time to market is practically direct for the research platform. As soon as the platform is available and validated within the project, it can be proposed to the research market. For the service robotics applications, we can expect one to three years time to market to bring the solutions developed within NIFTi. While the market size for research is difficult to estimate, it is clear that rescue applications, firefighting, and police applications represent high potential. This will have to be evaluated to propose a clear action plan towards to the market.

3.2.1.2 BLUE, DFKI: Exploitation of multi-modal human-robot interaction DFKI and BLUE will lead efforts for commercial exploitation of the results in multi-modal human-robot interaction for USAR missions (primarily WP3, with additional experience and results deriving from involvement in WPs 4 and 7). DFKI and BLUE envision to establish a joint venture, offering the HRI platform in conjunction with the rover technology as a *complete concept solution* to the USAR community. The exploitation will be based on a commercializable re-implementation of the platform developed within NIFTi. The NIFTi research software will remain publicly available, modulo the reservations stated in the consortium agreement.

DFKI and BLUE will prepare the first stage of a business plan by M12. This first business plan will serve three purposes. First, it will clearly outline the purpose of the joint venture. It will state the business goal ("commercial HRI for USAR"), and define the infrastructure for distribution, licensing, and development. Any other project partners to be involved will be identified at this point. Their involvement will be according to the mechanisms outlined in the consortium agreement on IPR. Second, the plan will present a detailed analysis of the market. The analysis will identify industrial, governmental, and non-profit non-governmental organizations in the USAR domain which would benefit from the technology to be marketed by the joint venture. The analysis will make these benefits explicit, as such and in relation to currently used technological practices in the domain. The analysis will couple the benefits to economical considerations, presenting a cost/benefit analysis given the funding patterns of different types of organizations in the USAR domain. The market analysis will be made available to the project as a PU-status deliverable in WP8 (dissemination).

Over the course of the project, DFKI and BLUE will actively establish contacts with potentially interested parties, e.g. through the NIFTi Rescue Days. DFKI and BLUE will update the business plan using the feedback from end user evaluations within NIFTi, and any feedback from the community through dissemination of the NIFTi project results.

DFKI and BLUE will prepare the second stage of the business plan by M44. This business plan will present an updated market- and benefit-analysis, and a financial analysis for starting up the joint venture. The financial analysis will include an analysis of funding sources, needed investments in capital equipment, an initial balance sheet, a break-even analysis, and income- and cash flow projections for the first 4 years. Depending on the financial climate for starting up joint ventures by project ending, the plan will envision a time-to-market of 18 to 24 months. The updated market analysis will again be provided as PU-status deliverable in WP8.

As a non-profit organization, DFKI cannot itself be directly involved in the joint venture. Instead, it will help establish a spin-off construction for the joint venture, (depending on feasibility given the then-current financial climate). DFKI has a long and successful track-record in establishing spin-off companies. Since 1998, DFKI has helped founding 50 spin-off companies. Several of the biggest success stories started within the Language Technology Lab (e.g. Xtramind, acrolynx).

Depending on the financial situation by the end of the project (notably with respect to establishing spin-offs in the ICT sector), and the feasibility of commercially exploiting the technology developed within NIFTi, DFKI and BLUE will reserve the right not to act on the (confidential) business plan.

References

[1] Belardinelli A., Pirri F., and Carbone A. *Attention In Cognitive Systems.*, volume LNAI 5395, chapter Motion saliency maps from spatiotemporal filtering., pages 112–123. Springer, NLD, 2009.

[2] C. Ackerman and L. Itti. Robot steering with spectral image information. *IEEE Transactions on Robotics*, 21(2):247–251, 2005.

[3] Natasha Alechina, Mehdi Dastani, Brian Logan, and John-Jules Ch. Meyer. A logic of agent programs. In *AAAI*, pages 795–800, 2007.

[4] I. Aslan, F. Xu, H. Uszkoreit, A. Krüger, and J. Steffen. Compass2008: Multimodal, multilingual and crosslingual interaction for mobile tourist guide applications. In *Proceedings of intelligent Technologies for interactive Entertainment (Intetain)*, 2005.

[5] J. Baldridge and G.-J. M. Kruijff. Multi-modal combinatory categorial grammmar. In *Proceedings of EACL'03*, Budapest, Hungary, 2003.

[6] J. Baldridge and G.J.M. Kruijff. Coupling CCG and hybrid logic dependency semantics. In *Proc. ACL 2002*, pages 319–326, Philadelphia, PA, 2002.

[7] J. Bares and Wettergreen D. Lessons from the development and deployment of dante ii. In *Proceedings of the 1997 Field and Service Robotics Conference*, 1997.

[8] T. Becker, C. Gerstenberger, I. Kruijff-Korbayová, A. Korthauer, M. Pinkal, M. Pitz, P. Poller, and J Schehl. Natural and intuitive multimodal dialogue for in-car applications: The sammie system. In *Proceedings of PAIS-2006*, 2006.

[9] A. Belardinelli and F. Pirri. A biologically plausible robot attention model, based on space and time. *Cognitive Processing*, pages 11–14, 2006.

[10] A. Belardinelli, F. Pirri, and A. Carbone. Robot task-driven attention. In *Proc. International Symposium on Practical Cognitive Agents & Robots (PCAR 06)*, pages 117–128, 2006.

[11] A. Belardinelli, F. Pirri, and A. Carbone. Robot task-driven attention. In *PCAR '06: Proceedings of the 2006 international symposium on Practical cognitive agents and robots*, New York, NY, USA, 2006. ACM Press.

[12] A. Belardinelli, F. Pirri, and A. Carbone. Spatial discrimination in task-driven attention, 2006.

[13] A. Belardinelli, F. Pirri, and A. Carbone. Spatial discrimination in task-driven attention. In *Proc. IEEE RO-MAN'06*, pages 321–327, Hatfield, UK, 2006.

[14] A. Belardinelli, F. Pirri, and A. Carbone. Bottom-up gaze shifts and fixations learning, by imitation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, pages 256–271, 2007.

[15] A. Belardinelli, F. Pirri, and A. Carbone. Gaze motion clustering in scan path estimation. *Cognitive Processing*, 9::, 2008. published online.

[16] A. Billard and R. Siegwart. Robot learning from demonstration. *Robotics and Autonomous Systems*, 47(2-3):65–67, 2004.

[17] P. Blackburn et al. Back and forth through time and events. In *Proceedings of the 9th Amsterdam Colloquium*, Amsterdam, The Netherlands, 1993.

[18] N. Blaylock, J. Allen, and G. Ferguson. Managing communicative intentions with collaborative problem solving. In J. van Kuppevelt and R.W. Smith, editors, *Current and New Directions in Discourse and Dialogue*, volume 22 of *Text, Speech and Language Technology*, pages 63–84. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[19] Stephen A. Block, Andreas F. Wehowsky, and Brian C. Williams. Robust execution on contingent, temporally flexible plans. In *AAAI*, 2006.

[20] D. Bonnafous, S. Lacroix, and T. Simeon. Motion generation for a rover on rough terrains. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2001.

[21] S Bouabdallah, M Becker, and R Siegwart. Autonomous miniature flying robots: Coming soon! *IEEE Robotics and Automation Magazine*, 13(3), September 2007.

[22] M. Brenner. Situation-aware interpretation, planning and execution of user commands by autonomous robots. In *Proceedings of IEEE RO-MAN '07*, Jeju, Korea, 2007.

[23] M. Brenner, N. Hawes, J. Kelleher, and J. Wyatt. Mediating between qualitative and quantitative representations for task-orientated human-robot interaction. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2072–2077, Hyderabad, India, January 2007.

[24] M. Brenner and B. Nebel. Continual planning and acting in dynamic multiagent environments. In *Proceedings of the International Symposium on Practical Cognitive Agents and Robots*, Perth, Australia, 2006.

[25] T. Brick and M. Scheutz. Incremental natural language processing for hri. In *Proceedings of the 2007 ACM Conference on Human-Robot Interaction (HRI'07)*, pages 263–270, 2007.

[26] C. Brooks, K. Iagnemma, and S. Dubowsky. Vibration-based terrain analysis for mobile robots. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005*, pages 3415–3420, 2005.

[27] Neil D. B. Bruce and John K. Tsotsos. An attentional framework for stereo vision. In *CRV*, pages 88–95, 2005.

[28] D. Bruemmer, D.A. Few, C. Nielse, and Walton M.C. Mixed-initiative control for collaborative countermine operations. In preparation, 2007.

[29] D. Bruemmer, J.L. Marble, D.D. Dudenhoeffer, M. Anderson, and M.D. McKay. Mixed-initiative control for remote characterization of hazardous environments. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 2003.

[30] P. Buschka and A. Saffiotti. Some notes on the use of hybrid maps for mobile robots. In *Proc. of the 8th Int. Conference on Intelligent Autonomous Systems*, Amsterdam, The Netherlands, March 2004.

[31] A. Carbone, D. Ciacelli, A Finzi, and F. Pirri. Autonomous attentive exploration in search and rescue scenarios. In *WAPCV*, pages 431–446, 2007.

[32] A. Carbone, A. Finzi, A. Orlandini, and F. Pirri. Model-based control architecture for attentive robots in rescue scenarios. *Autonomous Robots*, 28:87–120, 2008.

[33] M. Carpenter, T. Nagell, and M. Tomasello. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monographs of the Society for Research in Child Development, 63 (4, Serial No. 255) 1998.

[34] J.B. Carroll. Cognitive abilities: The state of the art. *Psychological Science*, 3:266–270, 1992.

[35] J. Casper. *Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center.* PhD thesis, University of South Florida, 2002.

[36] A. Castano, A. Talukder, and L. Matthies. Obstacle detection and terrain classification for autonomous off-road navigation. *Autonomous Robots*, 18:81–102, 2005.

[37] C. Castillo and C. Chang. A method to detect victims in search and rescue operations using template matching. *Safety, Security and Rescue Robotics, Workshop, 2005 IEEE International*, pages 201–206, 6-9 June 2005.

[38] S. Clark and J.R. Curran. Wide-coverage efficient statistical parsing with ccg and log-linear models. *Computational Linguistics*, 33(4):493–552, 2007.

[39] C. Coue, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere. Bayesian occupancy filtering for multitarget tracking: an automotive application. *Int. Journal of Robotics Research*, 25(1):19–30, January 2006.

[40] K.R. Coventry and S. Garrod. *Saying, Seeing and Acting. The Psychological Semantics of Spatial Prepositions.* Essays in Cognitive Psychology Series. Lawrence Erlbaum Associates, 2004.

[41] R. Dale and E. Reiter. Computational interpretations of the Gricean Maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263, 1995.

[42] C. Canudas de Wit, H. Khennouf, C. Samson, and O. Srdalen. *Nonlinear Control Design for Mobile Robots*, pages 121–156. Yuan F. Zheng, 1993.

[43] Yiannis Demiris and Bassam Khadhouri. Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems*, 54(5):361–369, 2006.

[44] D. DeVault and M. Stone. Domain inference in incremental interpretation. In *Proceedings of the Fourth Workshop on Inference in Computational Semantics (ICoS-4)*, 2003.

[45] Lowe D.G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[46] J.L. Drury, J. Scholtz, and H.A. Yanco. Awareness in human-robot interactions. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 912–918, 2003.

[47] J. Duncan. Disorganization of behaviour after frontal-lobe damage. *Cognitive Neuropsychology*, 3:271–290, 1986.

[48] Tobias Ehlgen and Tomáš Pajdla. Monitoring surrounding areas of truck-trailer combinations. In *ICVS 2006: Proceedings of the 5th International Conference on Vision Systems*, page 10, Bielefeld, Germany, March 2007. Applied Computer Science Group, Bielefeld University.

[49] M.R. Endsley. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting*, pages 97–101. Human Factors Society, 1988.

[50] M.R. Endsley. Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland, editors, *Situation awareness analysis and measurement*. Lawrence Erlbaum, 2000.

[51] M.R. Endsley and D.G. Jones. *Designing for Situation Awareness: An Approach to User-Centered Design*. Taylor & Francis, 2003.

[52] T Estier, R Piguet, R Eichhorn, and R Siegwart. Shrimp, a rover architecture for long range martian mission. In *Proc. of The Sixth ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA)*, December 2000.

[53] Y. Estier, B. Crausaz, M. Merminod, R. Lauria, R. Piguet, and R. Siegwart. An innovative space rover with extended climbing abilities. In *Proc. of the Space and Robotics*, 2000.

[54] Tara A. Estlin, Daniel Gaines, Caroline Chouinard, Rebecca Castano, Benjamin Bornstein, Michele Judd, Issa A. D. Nesnas, and Robert Anderson. Increased mars rover autonomy using ai planning, scheduling and execution. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 4911–4918. IEEE, 2007.

[55] R. Fay, U. Kaufmann, A. Knoblauch, H.r Markert, and G. Palm. Combining visual attention, object recognition and associative information processing in a neurobotic system. In *Biomimetic Neural Learning for Intelligent Robots*, pages 118–143, 2005.

[56] D. Ferguson and A. Stentz. Field d*: An interpolation-based path planner and replanner. In *Proceedings of the International Symposium on Robotics Research (ISRR)*, 2005.

[57] A. Finzi and F. Pirri. Representing flexible temporal behaviors in the situation calculus. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 436–441, 2005.

[58] T. Fong, C. Thorpe, and C. Baur. Collaboration, dialogue, and human-robot interaction. In *Proceedings of 10th International Symposium on Robotics Research*, 2001.

[59] T.W. Fong, I. Nourbakhsh, R. Ambrose, R. Simmons, A. Schultz, and J. Scholtz. The peer-to-peer human-robot interaction project. In *Proceedings of the AIAA Conference Space 2005*, 2005.

[60] T.W. Fong, C. Thorpe, and C. Baur. Robot, asker of questions. *Robotics and Autonomous Systems*, pages 235–243, 2003.

[61] S. Forster and N Lavie. Attentional capture by entirely irrelevant distractors. *Visual Cognition*, 16:200–214, 2008.

[62] Birte U. Forstmann, Marcel Brass, Iring Koch, and D. Yves Von Cramon. Voluntary selection of task sets revealed by functional magnetic resonance imaging. *J. Cognitive Neuroscience*, 18(3):388–398, 2006.

[63] Emily B. Fox, Erik B. Sudderth, Michael I. Jordan, and Alan S. Willsky. Nonparametric bayesian learning of switching linear dynamical systems. In *NIPS*, 2008.

[64] Vojtěch Franc and Bogdan Savchynskyy. Discriminative learning of max-sum classifiers. *Journal of Machine Learning Research*, 9:67–104, 2008.

[65] S. Frintrop, P. Jensfelt, and H. Christensen. Attentional landmark selection for visual slam. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06)*, 2006.

[66] S. Frintrop, E. Rome, A. Nüchter, and H. Surmann. A Bimodal Laser-Based Attention System. *Journal of Computer Vision and Image Understanding (CVIU)*, 100(1-2, Special Issue on Attention and Performance):124–151, Oct-Nov 2005. Class-AA.

[67] Sandra Clara Gadanho. Learning behavior-selection by emotions and cognition in a multi-goal robot task. *J. Mach. Learn. Res.*, 4:385–412, 2003.

[68] L. Getoor and B. Taskar. *Statistical Relational Learning*. The MIT Press, 2007.

[69] M. Ghallab and H. Laruelle. Representation and control in ixtet, a temporal planner. In *AIPS 1994*, pages 61–67, 1994.

[70] N.A. Hawes G.J.M. Kruijff, M. Brenner. Continual planning for cross-modal situated clarification in human-robot interaction. In *Proceedings of the 17th International Symposium on Robot and Human Interactive Communication (RO-MAN 2008)*. IEEE, 8 2008.

[71] L. Goncalves, C. Distante, A. Oliveira, D. Wheeler, and R. Grupen. Neural mechanisms for learning of attention control and pattern categorization as basis for robot cognition. In *In Proc. (IROS 2000)*, pages 1088–1093, 2000.

[72] M. Goodrich, D. Olsen, J. Crandall, and T. Palmer. Experiments in adjustable autonomy. In *Proceedings of IJCAI Workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, 2001.

[73] M. A. Goodrich, B. S. Morse, D. Gerhardt, J. L. Cooper, M. Quigley, J. A. Adams, and C. Humphrey. Supporting wilderness search and rescue using a camera-equipped mini uav. *Journal of Field Robotics*, 25(1-2):89–110, 2008.

[74] M.A. Goodrich. Using Models of Cognition in HRI Evaluation and Design. In *Proceedings of the AAAI 2004 Fall Symposium Series: The Intersection of Cognitive Science and Robotics: From Interfaces to Intelligence*, pages 21–24, 2004.

[75] T. Grenager, R. Powers, and Y. Shoham. Dispersion games: General definitions and some specific learning results. In *AAAI/IAAI*, pages 398–403, 2002.

[76] M. Grootjen, M.A. Neerincx, K.D. Stolk, J.C.M. van Weert, and E.P.B. Bierman. Design and user evaluation of an interface prototype that adapts to the operator's cognitive task load. In *Proceedings of the 4th International Conference of the Augmented Cognition.*, 2007.

[77] M. Grootjen, M.A. Neerincx, and J.A. Veltman. Cognitive task load in naval ship control centres: from identification to prediction. *Ergonomics*, 49:1238–1264, 2006.

[78] M. Grootjen, M.A. Neerincx, JCM Weert, and K.P. Truong. Measuring Cognitive Task Load on a Naval Ship: Implications of a Real World Environment. In *Proceedings of International Conference on Human-Computer Interaction (HCII'07)*. Springer, 2007.

[79] B.J. Grosz and L. Hunsberger. The dynamics of intention in collaborative activity. *Cognitive Systems Research*, 7(2-3):259–272, 2006.

[80] B.J. Grosz and S. Kraus. The evolution of sharedplans. In A. Rao and M. Wooldridge, editors, *Foundations and Theories of Rational Agency*, pages 227–262. Springer, 1999.

[81] Slawomir Grzonka, Giorgio Grisetti, and Wolfram Burgard. Towards a navigation system for autonomous indoor flying. In *Proc. IEEE International Conference on Robotics and Automation (ICRA'09)*, Kobe, Japan, 2009.

[82] J.H.L. Hansen and S.A. Patil. Speech under stress: Analysis, modeling and recognition. In C. Müller, editor, *Speaker Classification I: Fundamentals, Features, and Methods*, volume 4343 of *Lecture Notes In Artificial Intelligence*, pages 108–137. Springer-Verlag, Berlin, 2007.

[83] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[84] M. Havlena, T. Pajdla, and K. Cornelis. Structure from omnidirectional stereo rig motion for city modeling. In *VISAPP 2008 - International Conference on Computer Vision Theory and Applications*. INSTICC - Institute for Systems and Technologies of Information, Control and Communication, Setubal, Portugal, January 2008.

[85] Michal Havlena, Kurt Cornelis, and Tomáš Pajdla. Towards city modeling from omnidirectional video. In Michael Grabner and Helmut Grabner, editors, *CVWW 2007: Proceedings of the 12th Computer Vision Winter Workshop*, pages 123–130, Graz, Austria, February 2007. Institute for Computer Graphics and Vision, Graz University of Technology, Gratz, Austria, Verlag der Technischen Universität Graz.

[86] N Hawes, A. Sloman, J. Wyatt, M. Zillich, H. Jacobsson, G.J.M Kruijff, M. Brenner, G. Berginc, and D. Skocaj. Towards an integrated robot with multiple cognitive functions. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI-07)*, 2007.

[87] J. Henderson, O. Lemon, and K. Georgila. Hybrid reinforcement / supervised learning of dialogue policies from fixed datasets. *Computational Linguistics*, 34(4), 2008.

[88] Mathieu Hillion and Florent Lamiraux. Taking into account velocity and acceleration bounds in nonholonomic trajectory deformation. In *ICRA*, pages 3080–3085, 2007.

[89] Stephen C. Hirtle and John Jonides. Evidence for hierarchies in cognitive maps. *Memory and Cognition*, 13:208–217, 1985.

[90] B.A. Hockey, O. Lemon, E. Campana, L. Hiatt, G. Aist, J. Hieronymus, A. Gruenstein, and J. Dowding. Targeted help for spoken dialogue systems: intelligent feedback improves naive users' performance. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*, pages 147 – 154, Budapest, Hungary, 2003.

[91] Thomas Howard and Alonzo Kelly. Optimal rough terrain trajectory generation for wheeled mobile robots. *International Journal of Robotics Research*, 26(1):141–166, March 2007.

[92] Celine Hudelot, Nicolas Maillot, and Monique Thonnat. Symbol grounding for semantic image interpretation: From image data to semantics. In *ICCVW '05: Proceedings of the Tenth IEEE International Conference on Computer Vision Workshops*, page 1875, Washington, DC, USA, 2005. IEEE Computer Society.

[93] H. Hügli, T. Jost, and N. Ouerhani. Model performance for visual attention in real 3d color scenes. In *IWINAC-2005*, pages 469–478, 2005.

[94] Fraunhofer IAIS. Technical data of the volksbot. In *http://www.volksbot.de*, 2008.

[95] G. Indiveri. Kinematic Time-invariant Control of a 2D Nonholonomic Vehicle. In *Proceedings of the 38th Conference on Decision and Control, (CDC '99)*, pages 2112–2117, Phoenix, USA, December 1999.

[96] G. Indiveri and M. L. Corradini. Switching linear path following for bounded curvature car-like vehicles. In *In Proceedings of the IFAC Symposium on Intelligent Autonomous Vehicles (IAV 04)*, 2004.

[97] G. Indiveri, A. Nuechtr, and K. Lingemann. High speed differential drive mobile robot path following control with bounded wheel speed commands. In *Proceedings of the IEEE International Conference Robotics and Automation (ICRA '07)*, pages 2202 – 2207, 2007.

[98] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, Mar 2001.

[99] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, Mar 2001.

[100] H. Jacobsson, N. Hawes, G.J.M. Kruijff, and J. Wyatt. Crossmodal content binding in information-processing architectures. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI'08)*, Amsterdam, The Netherlands, March 2008.

[101] A. Jacoff and E. Weiss, B. Messina. Performance metrics and test arenas for autonomous mobile robots. In *http://www.isd.mel.nist.gov/projects/USAR/*, 2007.

[102] A. Jameson. Modeling both the context and the user. *Personal and Ubiquitous Computing*, 5(1):29–33, 2004.

[103] A. Jameson. Adaptive interfaces and agents. In J. A. Jacko and A. Sears, editors, *Human-computer interaction handbook*. Erlbaum, 2006.

[104] A. Jameson, J. Kiefer, C. MÏler, F. Wittig, and R. Rummer. Assessment of a user's time pressure and cognitive load on the basis of features of speech. In M. Crocker and J. Siekmann, editors, *Resource-adaptive cognitive systems*. Springer Verlag, 2008.

[105] C. Jansen, S. de Vries, and M. Duistermaat. Optimizing the presentation of uav images in an attack helicopter cockpit. In *Proceedings of the Human factors and ergonomics society (HFES) 50th annual meeting*, pages 131–135, 2006.

[106] T. Jebara and A. Pentland. Action reaction learning: Automatic visual analysis and synthesis of interactive behaviour. In *ICVS*, pages 273–292, 1999.

[107] M.J. Johnson, K.Jr. Intlekofer, H.H. Jung, J.M. Bradshaw, J. Allen, N. Suri, and M. Carvalho. Coordinated operations in mixed teams of humans and robots. In *Proceedings of the 2008 IEEE International Conference on Distributed Human-Machine Systems (DHMS 2008)*, pages 63–68, Athens, Greece, March 2008.

[108] M. Johnston and S. Bangalore. Finite-state multimodal integration and understanding. *Journal of Natural Language Engineering*, 11(2):159–187, 2005.

[109] M. Johnston and S. Bangalore. Learning edit machines for robust multimodal understanding. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006.

[110] A.K. Jonsson, P.H. Morris, N. Muscettola, K. Rajan, and B.D. Smith. Planning in interplanetary space: Theory and practice. In *Artificial Intelligence Planning Systems*, pages 177–186, 2000.

[111] F. Jurie and M. Dhome. Real time robust template matching. In *British Machine Vision Conference*, pages 123–131, 2002.

[112] R. Kaestner, S. Thrun, M. Montemerlo, and M. Whalley. A non-rigid approach to scan alignment and change detection using range sensor data. In *Proceedings of the 5th International Conference on Field and Service Robotics*, 2005.

[113] S. Kasderidis and J. G. Taylor. Attention-based robot control. In *KES*, pages 615–621, 2003.

[114] Kazuhiko Kawamura, Tamara E. Rogers, and Xinyu Ao. Development of a cognitive model of humans in a multi-agent framework for human-robot interaction. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 1379–1386, New York, NY, USA, 2002. ACM.

[115] J.D. Kelleher, F. Costello, and J.A. van Genabith. Dynamically updating and interrelating representations of visual and linguistic discourse. *Artificial Intelligence*, 67(1-2):62–102, 2005.

[116] J.D. Kelleher and G.J.M. Kruijff. Incremental generation of spatial referring expressions in situated dialog. In *Proceedings of ACL/COLING 2006*, Sydney, Australia, 2006.

[117] A. L. Kemurdjian, V. Gromov, V. Mishkinyuk, V. Kucherenko, and P. Sologub. Small marsokhod configuration. In *International Conference on Robotics & Automation*, Nice, 1992.

[118] P. Knoeferle and M.C. Crocker. The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science*, 2006.

[119] N. Koenig, S. Chernova, C. Jones, M. Loper, and O. C. Jenkins. Hands-free interaction for human-robot teams. In G.J.M. Kruijff, H. Zender, M. Hanheide, and B. Wrede, editors, *Proceedings of the ICRA 2008 workshop Social Interaction with Intelligent Indoor Robots*, Pasadena CA, 2008.

[120] M. Koes, I. R. Nourbakhsh, and K. P. Sycara. Heterogeneous multirobot coordination with spatial and temporal constraints. In *AAAI*, pages 1292–1297, 2005.

[121] M. Koes, K. Sycara, and I.R. Nourbakhsh. A constraint optimization framework for fractured robot teams. In *AAMAS*, pages 491–493, 2006.

[122] S Kolski. *Mobile Robots: Perception and Mobile Robots: Perception and Navigation.* International Journal of Advanced Robotic Systems. Pro Literatur Verlag, 2007.

[123] J. Zico Kolter, Youngjun Kim, and Y. Ng. Andrew. Stereo vision and terrain modeling for quadruped robots. In *Proceedings of the International Conference on Robotics and Automation*, 2009.

[124] Filip Korč and Václav Hlaváč. *Detection and Tracking of Humans in Single View Sequences Using 2D Articulated Model*, volume 36 of *Computational Imaging and Vision*, chapter 5, pages 105–130. Springer Verlag, Heidelberg, Germany, 1 edition, 2007.

[125] E. Krahmer and M Swerts. Displayed, but not felt: Production and perception of congruent and incongruent emotional expressions. accepted, 2008.

[126] E. Krahmer and M. Theune. Efficient context-sensitive generation of referring expressions. In K. van Deemter and R.Kibble, editors, *Information Sharing: Givenness and Newness in Language Processing*. CSLI Publications, Stanford, CA, USA, 2002.

[127] H.U. Krieger. Where temporal description logics fail: Representing temporally-changing relationships. In A. Dengel, K. Berns, T.M. Breuel, F. Bomarius, and T.R. Roth-Berghofer, editors, *KI 2008: Advances in Artificial Intelligence*, number 5243 in Lecture Notes in Artificial Intelligence, pages 249–257. Springer, 2008.

[128] H.U. Krieger, B. Kiefer, and T. Declerck. A framework for temporal representation and reasoning in business intelligence applications. In *Proceedings of the AAAI 2008 Spring Symposium on AI Meets Business Rules and Process Management*, Stanford, CA, 2008.

[129] G.J.M. Kruijff. Context-sensitive utterance planning for ccg. In *Proceedings of the European Workshop on Natural Language Generation*, Aberdeen, Scotland, 2005.

[130] G.J.M. Kruijff and M. Brenner. Modelling spatio-temporal comprehension in situated human-robot dialogue as reasoning about intentions and plans. In *Proceedings of the AAAI 2007 Spring Symposium on Intentions in Intelligent Systems*, 2007.

[131] G.J.M Kruijff, P. Lison, T. Benjamin, H. Jacobsson, H. Zender, and I. Kruijff-Korbayová. Situated dialogue processing for human-robot interaction. In H.I.C Christensen, G.J.M. Kruijff, and J.L. Wyatt, editors, *Cognitive Systems*. Springer Verlag, 2009.

[132] G.J.M. Kruijff, H. Zender, P. Jensfelt, and H.I. Christensen. Clarification dialogues in human-augmented mapping. In *Proceedings of the 1st Annual Conference on Human-Robot Interaction (HRI'06)*, 2006.

[133] G.J.M. Kruijff, H. Zender, P. Jensfelt, and H.I. Christensen. Situated dialogue and spatial organization: What, where... and why? *International Journal of Advanced Robotic Systems*, 4(2):125–138, 2007.

[134] A.A. Kruse and D.D. Schmorrow. Session overview: Foundations of augmented cognition. *Foundations of Augmented Cognition*, pages 441–445, 2005.

[135] T. Kubota, Y. Kuroda, Y. Kunii, and I. Natakani. Micro planetary rover micro5. In *Proceedings of Fifth International Symposium on Artificial Intelligence, Robotics and Automation in Space (ESA SP-440)*, pages 373–378, Noordwijk, 1999.

[136] Lionel Lapierre, Rene Zapata, and Pascal Lepinay. Combined path-following and obstacle avoidance control of a wheeled robot. *Int. J. Rob. Res.*, 26(4):361–375, 2007.

[137] S. Larsson. *Issue-Based Dialogue Management*. Phd thesis, Department of Linguistics, Göteborg University, Göteborg, Sweden, 2002.

[138] M. Lauria, F. Conti, P.-A. Maeusli, M. Van Winnendael, R. Bertrand, and R. Siegwart. Design and control of an innovative micro-rover. In *Proceedings of 5th ESA Workshop on Advanced Space Technologies for Robotics and Automation*, Noordwijk, 1998.

[139] M. Lauria, Y. Piguet, and R. Siegwart. Octopus - an autonomous wheeled climbing robot. In *Proceedings of the Fifth International Conference on Climbing and Walking Robots*, Bury St Edmunds and London, UK, 2002.

[140] D. Lecking, O. Wulf, and B. Wagner. Localization in a wide range of industrial environments using relative 3d ceiling features. In *International Conference on Emerging Technologies and Factory Automation*, 2008.

[141] O. Lemon. Adaptive natural language generation in dialogue using reinforcement learning. In *Proceedings of the Workshop Series on the Semantics and Pragmatics of Dialogue (LONDIAL 2008)*, 2008.

[142] I. Leppanen, S. Salmi, and A. Halme. Workpartner hut automation's new hybrid walking machine. In *CLAWAR'98 First international symposium*, Brussels, 1998.

[143] K. Lerman, C. Jones, A. Galstyan, and M.J. Mataric. Analysis of dynamic task allocation in multi-robot systems. *I. J. Robotic Res.*, 25(3):225–241, 2006.

[144] Y. Lespérance and H. J. Levesque. Indexical knowledge and robot action - a logical account. *Artificial Intelligence*, 73:69–115, 1995.

[145] Hector Levesque and Gerhard Lakemeyer. *Cognitive Robotics*, incollection 24. Cognitive Robotics. Handbook of Knowledge Representation. Elsevier, 2007. Editor: Frank van Harmelen and Vladimir Lifschitz and Bruce Porter.

[146] Zhe Lin, Larry S. Davis, David Doermann, and Daniel DeMenthon. Hierarchical part-template matching for human detection and segmentation. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 14-21 Oct. 2007.

[147] P. Lison and G.-J. M. Kruijff. Salience-driven contextual priming of speech recognition for human-robot interaction. In *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008)*, Patras, Greece, 2008.

[148] P. Lison and G.J. M. Kruijff. An integrated approach to robust processing of situated spoken dialogue. In *Proceedings of the International Workshop on Semantic Representation of Spoken Language (SRSL'09)*, Athens, Greece, 2009.

[149] K. Lochbaum, B.J. Grosz, and C.L. Sidner. Discourse structure and intention recognition. In R. Dale, H. Moisl, , and H. Somers, editors, *A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text*. Marcel Dekker, New York, 1999.

[150] R. Looije, F. Cnossen, and MA Neerincx. Incorporating guidelines for health assistance into a socially intelligent robot. In *ROMAN 2006*, pages 515–520, 2006.

[151] M. Maguire. Methods to support human-centred design. *International Journal of Human-Computer Studies*, 55(4):587–634, 2001.

[152] A Martinelli, V Nguyen, N Tomatis, and R Siegwart. A relative map approach to slam based on shift and rotation invariants. *Robotics and Autonomous Systems*, 2006.

[153] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In P.L. Rosin and D. Marshall, editors, *Proceedings of the British Machine Vision Conference*, volume 1, pages 384–393, London, UK, September 2002. BMVA.

[154] Jiří Matas, Ondřej Chum, Martin Urban, and Tomáš Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, September 2004.

[155] Jiří Matas, Karel Zimmermann, Tomáš Svoboda, and Adrian Hilton. Learning efficient linear predictors for motion estimation. In Subhashis Banerjee Rangachar Kasturi, editor, *Proceedings of 5th Indian Conference on Computer Vision, Graphics and Image Processing*, LNCS4338, pages 445–456, December 2006.

[156] PAB Merkx, K.P. Truong, and M.A. Neerincx. Inducing and measuring emotion through a multi-player first-person shooter computer game. *Proceedings of the Computer Games Workshop*, 2007.

[157] M. Michalowski, S. sabanovic, and H. Kozima. A dancing robot for rhythmic social interaction. In *Proceedings of ACM Conference on Human-Robot Interaction*, pages 89–96, 2007.

[158] Branislav Mičušík and Tomăš Pajdla. Structure from motion with wide circular field of view cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1135–1149, July 2006.

[159] K. Mikolajczyk, T Tuytelaars, C Schmid, A Zisserman, J Matas, F Schaffalitzky, T Kadir, and L van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(7):43 – 72, November 2005.

[160] Krystian Mikolajczyk and Jiří Matas. Improving sift for fast tree matching by optimal linear projection. In Dimitris Metaxas, Baba Vemuri, Amnon Shashua, and Harry Shum, editors, *ICCV 2007: Proceedings of Eleventh IEEE International Conference on Computer Vision*, page 8, Los Alamitos, USA, October 2007. IEEE Computer Society, IEEE Computer Society Press.

[161] Krystian Mikolajczyk, Cordelia Schmid, and Andrew Zisserman. Human detection based on a probabilistic assembly of robust part detectors. In *European Conference on Computer Vision*, volume I, pages 69–81, 2004.

[162] B. Milch and S. J. Russell. General-purpose mcmc inference over relational structures. In *UAI*, 2006.

[163] Brian Milch and Stuart J. Russell. First-order probabilistic languages: Into the unknown. In *ILP*, pages 10–24, 2006.

[164] C.A. Miller and R. Parasuraman. Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(1):57–75, 2007.

[165] M. Moens and M. Steedman. Temporal ontology and temporal reference. *Journal of Computational Linguistics*, 14:15–28, 1988.

[166] Benoit Morisset and Malik Ghallab. Learning how to combine sensory-motor functions into a robust behavior. *Artif. Intell.*, 172(4-5):392–412, 2008.

[167] M.C. Mozer and S.P. Vecera. Object- and space-based attention. In L. Itti, G. Rees, and J. Tsotsos, editors, *The encyclopedia of the neurobiology of attention*, pages 130–134. Elsevier Press, 2005.

[168] O. Martinez Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard. Supervised semantic labeling of places using information extracted from sensor data. *Robotics and Autonomous Systems*, 55(5):391–402, 2007.

[169] C. Mueller. Classifying speakers according to age and gender. In C. Müller, editor, *Speaker Classification II*, Lecture Notes In Artificial Intelligence. Springer-Verlag, Berlin, 2007.

[170] J.M.F. Mulder. Supporting the internal battle. In *Thirteenth International Ship Control Systems Symposium (SCSS)*, 2003.

[171] C. Müller, F. Wittig, and J. Baus. Exploiting speech for recognizing elderly users to respond to their special needs. In *Proceedings of EuroSpeech 2003*, pages 1305–1308, Geneva, Switzerland, 2003.

[172] R.R. Murphy. Human-robot interaction in rescue robotics. *Systems, Man and Cybernetics Part C: Applications and Reviews*, 34(2):138– 153, 2004.

[173] R.R. Murphy and J.L. Burke. Up from the rubble: Lessons learned about hri from search and rescue. In *Proceedings of the 49th Annual Meetings of the Human Factors and Ergonomics Society*, 2005.

[174] G. Gruener N. Ouerhani, H. Hugli and A. Codourey. A visual attention-based approach for automatic landmark selection and recognition. In *Lecture Notes in Computer Science*, volume 3368/2005, pages 183–195. Springer Berlin / Heidelberg, 2005.

[175] M. A. Neerincx. *Cognitive task load analysis: allocating tasks and designing support*, chapter 13, pages 283–305. Lawrence Erlbaum, 2003.

[176] M.A. Neerincx. Modelling Cognitive and Affective Load for the Design of Human-Machine Collaboration. *LECTURE NOTES IN COMPUTER SCIENCE*, 4562:568, 2007.

[177] M.A. Neerincx, S. Kennedie, F. Grootjen, and M. Grootjen. Modelling cognitive task load and emotion for adaptive maritime interfaces. In *Proceedings of the 5th International Conference of the Augmented Cognition.*, San Diego, CA, USA, 19-24 July 2009.

[178] M.A. Neerincx, S. Kennedie, F. Grootjen, and M. Grootjen. Modelling Cognitive Task Load and Emotion for Adaptive Maritime Interfaces. *Proceedings of the 5th International Conference of the Augmented Cognition*, page 10, 2009.

[179] M.A. Neerincx and J. Lindenberg. *Situated cognitive engineering for complex task environments*, chapter Situated cognitive engineering for complex task environments, page coming soon. Aldershot, UK: Ashgate Publishing Limited, 2008. Editor: Schraagen, J.M., Militello, L., Ormerod, T., & Lipshitz, R.

[180] M.A. Neerincx, J. Lindenberg, and M. Grootjen. Accessibility on the Job: Cognitive Capacity Driven Personalization. In *HCII 2005, Volume 7 - Universal Access in HCI: Exploring New Interaction Environments*, page 10 pages. St. Louis, Missouri: MIRA Digital Publishing, 2005.

[181] M.A Neerincx, J. Lindenberg, N. Smets, T. Grant, A. Bos, U. Brauer, and M. Wolff. Cognitive Engineering for Long Duration Missions: Human-Machine Collaboration on the Moon and Mars. In *Second IEEE International Conference on Space Mission Challenges for Information Technology, 2006.*, pages 40–46, 2006.

[182] M.A. Neerincx, G.M. Te Brake, J.G.M. Van de Ven, H.F.R. Arciszewski, T.E. De Greef, and J. Lindenberg. Situated cognitive engineering: Developing adaptive track handling support for naval command and control centers. In *Proceedings of Human-Centered Processes*, 2008.

[183] A. Newell. *Unified theories of cognition*. Harvard University Press, 1990.

[184] V Nguyen and R Siegwart. Information relative map: Going toward constant time slam. In *Proc. of The European Robotics Symposium (EUROS)*, 2008.

[185] E. Niebur, L. Itti, and C. Koch. Controlling the focus of visual selective attention. In L. Van Hemmen, E. Domany, and J. Cowan, editors, *Models of Neural Networks IV: Early Vision and Attention*, pages 247–276. Springer Verlag, Aug 2001.

[186] C. W. Nielsen, M. A. Goodrich, , and B. Ricks. Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics and Automation*, 23(5):927–941, October 2007.

[187] Y. Niwa and E. Hollnagel. Enhancing Operator Control by Adaptive Alarm Presentation. *International Journal of Cognitive Ergonomics*, 5(3):367–384, 2001.

[188] D. A. Norman and T. Shallice. *Consciousness and Self-Regulation: Advances in Research and Theory*, volume 4, chapter Attention to action: Willed and automatic control of behaviour. Plenum Press, 1986. Editor: Davidson, R. J. and Schwartz, G. E. and Shapiro, D.

[189] A. Nüchter, K. Lingemann, J. Hertzberg, and H. Surmann. 6d slam – 3d mapping outdoor environments: Research articles. *Journal of Field Robotics*, 24(8-9):699–722, 2007.

[190] A. Nüchter, K. Lingemann, J. Hertzberg, H. Surmann, K. Pervölz, M. Hennig, K.R. Tiruchinapalli, R. Worst, and T. Christaller. Mapping in rescue environments with kurt3d (best paper awarded after vizeworld champoins 2004). In *Proceedings of the IEEE International Workshop on Safty, Security and Rescue Robotics 2005 (SSRR '05)*, pages 158–163, 2005. Conf-A, Best paper award.

[191] A. Nüchter, K. Lingemann, J. Hertzberg, O. Wulf, B. Wagner, and H. Surmann. 3d mapping with semantic knowledge. In *RoboCup International Symposium 2005: Robot Soccer World Cup IX*, pages 335–346, Osaka, Japan, 2005. Conf-A.

[192] A. Nüchter, H. Surmann, and J. Hertzberg. Automatic classification of objects in 3d laser range scans. In *Proc. 8th Conf. on Intelligent Autonomous Systems*, pages 963–970, 2004. Conf-A.

[193] A. Nüchter, H. Surmann, K. Lingemann, and J. Hertzberg. Semantic Scene Analysis of Scanned 3D Indoor Environments. In *Proceedings of the of the 8th International Fall Workshop Vision, Modeling, and Visualization (VMV '03)*, pages 215 – 222, Munich, Germany, November 2003.

[194] T. Pajdla and V. Hlaváč. Image-based self-localization by means of zero phase representation in panoramic images. In S. Singh, N. Murshed, and W. Kropatsch, editors, *Advances of Pattern Recognition, Proceedings of the 2nd International Conference on Advanced Pattern Recognition*, volume 2013 of *Lecture Notes in Computer Science*, pages 24–33, Heidelberg, Germany, March 2001. IAPR, Springer-Verlag.

[195] I. Paraboni, K. van Deemter, and J. Masthoff. Generating referring expressions: Making referents easy to identify. *Computational Linguistics*, 33(2):229–254, June 2007.

[196] R. Parasuraman, T.B. Sheridan, and C.D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 30(3):286–297, 2000.

[197] S.A. Patil and J.H.L. Hansen. Detection of speech under physical stress: Model development, sensor selection, and feature fusion. In *Proceedings of INTERSPEECH 2008*, 2008.

[198] R. Philippsen. A light formulation of the e* interpolated path replanner. Technical report, Autonomous Systems Lab, Ecole Polytechnique Federale de Lausanne, 2006.

[199] F. Pirri, A. Carbone, and A. Belardinelli. Patent number rm2007a000526: Gaze machine, 2007.

[200] F. Pirri and A. Pascucci. Implicit and explicit analysis of shapes. In *ICVW 2006 - second international cognitive vision workshop*, 2006.

[201] Fiora Pirri and Raymond Reiter. Some contributions to the metatheory of the situation calculus. *J. ACM*, 46(3):325–361, 1999.

[202] D. Poole. The independent choice logic and beyond. In *Probabilistic Inductive Logic Programming*, pages 222–243, 2008.

[203] M.I. Posner. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32-A:3–25, 1980.

[204] C.J. van Dongen P.P. van Maanen, L. de Koning. Design and validation of habta: Human attention-based task allocator. In *Proceedings of the First International Workshop on Human Aspects in Ambient Intelligenc*, 2007.

[205] D. Prusa and V. Hlavac. Mathematical formulae recognition using 2d grammars. In *ICDAR '07: Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2*, pages 849–853, Washington, DC, USA, 2007. IEEE Computer Society.

[206] M. Purver. *The Theory and Use of Clarification Requests in Dialogue.* PhD thesis, King's College, University of London, 2004.

[207] Z. Rabinovich and J.S. Rosenschein. Multiagent coordination by extended markov tracking. In *AAMAS*, pages 431–438, 2005.

[208] J. Rasmussen. *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering.* Elsevier Science Inc. New York, NY, USA, 1986.

[209] R. Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems.* MIT Press, Cambridge, MA, 2001.

[210] N. Reithinger, J. Alexandersson, T. Becker, A. Blocher, R. Engel, M. Löckelt, J. Müller, N. Pfleger, P. Poller, M. Streit, and V. Tschernomas. Smartkom - adaptive and flexible multimodal access to multiple applications. In *Proceedings of ICMI 2003*, Vancouver, B.C., 11 2003.

[211] V. Rieser and O. Lemon. Using machine learning to explore human multimodal clarification strategies. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, Sydney, 2006.

[212] V. Rieser and O. Lemon. Learning effective multimodal dialogue strategies from wizard-of-oz data: Bootstrapping and evaluation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL/HLT)*, Columbus, Ohio, 2008.

[213] V. Rieser and O. Lemon. Simulation-based learning of optimal multimodal presentation strategies from wizard-of-oz data. In *Proceedings of the AISB Symposium on Multimodal Output Generation (MOG)*, Aberdeen, 2008.

[214] D. Roy. Grounding words in perception and action: Computational insights. *Trends in Cognitive Sciences*, 9(8):389–96, 2005.

[215] D.K. Roy. A computational model of three facets of meaning. In M. de Vega, A. Glenberg, and A. Graesser, editors, *Symbols, Embodiment, and Meaning.* Oxford University Press, 2008.

[216] Marra S. and Pirri F. Eyes and cameras calibration for 3d world gaze detection. In *International Conference on Computer Vision (ICVS)*, volume LNCS, pages 216–227, NLD, May 2008. Springer.

[217] S. Sanner and C. Boutilier. Practical linear value-approximation techniques for first-order mdps. In *UAI*, 2006.

[218] R. Sara. Robust correspondence recognition for computer vision. In Alfredo Rizzi and Maurizio Vichi, editors, *COMPSTAT 2006: Proceedings in Computational Statistics of 17th ERS-IASC Symposium*, pages 119–131, Heidelberg, Germany, August/September 2006. European Regional Section of the International Association for Statistical Computing, Physica-Verlag.

[219] G. Schindler and F. Dellaert. Inferring temporal order of images from 3d structure. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[220] Michail I. Schlesinger and Václav Hlaváč. *Ten lectures on statistical and structural pattern recognition*, volume 24 of *Computational Imaging and Vision*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[221] S. Schötz. Acoustic analysis of adult speaker age. In C. Müller, editor, *Speaker Classification I: Fundamentals, Features, and Methods*, volume 4343 of *Lecture Notes In Artificial Intelligence*, pages 88–107. Springer-Verlag, Berlin, 2007.

[222] M. Schroeder. Expressing degree of activation in synthetic speech. *IEEE transactions on audio, speech and language processing*, 14(4):1128–1136, 2006.

[223] B.P. Sellner and R. Simmons. Towards proactive replanning for multi-robot teams. In *Proceedings of the 5th International Workshop on Planning and Scheduling in Space 2006*, October 2006.

[224] M.P. Shanahan. A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition*, 15:433–449, 2006.

[225] Alexander Shekhovtsov, Ivan Kovtun, and Václav Hlaváč. Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, 112(1):91–99, October 2008.

[226] R. Siegwart, P. Lamon, T. Estier, M. Lauria, and R. Piguet. Innovative design for wheeled locomotion in rough terrain. *Journal of Robotics and Autonomous Systems, Elsevier*, 40(2):151–162, 2002.

[227] D.J. Simons and D.T. Levin. Change blindness. *Trends in Cognitive Sciences*, 1:261–267, 1997.

[228] S. Singh, R. Simmons, T. Smith, A. Stentz, Verma V., Yahja A., and K. Schwehr. Recent progress in local and global traversability for planetary rovers. In *IEEE International Conference on Robotics and Automation*, San Francisco, USA., 2000.

[229] S.Li, B. Wrede, and G. Sagerer. A computational model of multi-modal grounding. In *Proceedings of the ACL SIGdial workshop on discourse and dialog (SIGDIAL'06)*, pages 153–160, 2006.

[230] M. Šonka, V. Hlaváč, and R.D. Boyle. *Image Processing, Analysis and Machine Vision.* Thomson, Toronto, Canada, 3 edition, April 2007.

[231] L. Spinello, R. Triebel, and R. Siegwart. Multimodal people detection and tracking in crowded scenes. In *Proceedings of the AAAI Conference on Artificial Intelligence, Physically Grounded AI Special Track*, 2008.

[232] C. Stauffer and W.E.L. Grimson. Adaptive background models for real-time tracking. In *Proceedings of the Conference on Computer Vision and Pattern Recogntion*, volume 2, pages 2242–2252. IEEE Computer Society, June 1999.

[233] L. Steels. The symbol grounding problem has been solved. so what's next? In M. De Vega, G. Glennberg, and G. Graesser, editors, *Symbols, embodiment and meaning.* Academic Press, New Haven, 2008.

[234] L. Steels. Semiotic dynamics for embodied agents. *IEEE Intelligent Systems*, 21:32–38, 3.

[235] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. Common metrics for human-robot interaction. In *HRI 2006: 2006 ACM Conference on Human-Robot Interaction*, pages 33–40, Salt Lake City, Utah, 2006.

[236] H. W. Stone. Mars pathfinder microrover: A low-cost, low-power spacecraft. In *Proceedings of the 1996 AIAA Forum on Advanced Developments in Space Robotics*, Madison WI, 1996.

[237] J.W. Streefkerk, M.P. van Esch-Bussemakers, M.A. Neerincx, and R. Looije. Evaluating context-aware mobile user interfaces. *Handbook of Research on User Interface Design and Evaluation for Mobile Technology.*, pages 759–779, 2008.

[238] J.W. Streefkerk, C. Wiering, M.P. van Esch-Bussemakers, and M.A. Neerincx. Effects of presentation modality on team awareness and choice accuracy in a simulated police team task. In *Proceeding of the Human Factors and Ergonomics Society's 52nd Annual Meeting, New York, USA*, 2008.

Approved by EC on 2009-12-09

[239] H. Surmann, D. Holz, S. Blumenthal, T. Linder, P. Molitor, and V. Tretyakov. Teleoperated Visual Inspection and Surveillance with Unmanned Ground and Aerial Vehicles. *International Journal of Online Engineering 2008 (iJOE 2008)*, 4(4):26–38, Nov. 2008. ISSN: 1861-2121.

[240] H. Surmann, A. Nüchter, and J. Hertzberg. An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, 45(3-4):181–198, 2003.

[241] Hartmut Surmann, Ansgar Bredenfeld, Thomas Christaller, Reiner Frings, Ulrike Petersen, and Thomas Wisspeintner. The volksbot. In Emanuele Menegatti, editor, *Workshop Proceedings of Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS (SIMPAR)*, pages 551–561, November 2008. ISBN 978-88-95872-01-8.

[242] Tomáš Svoboda, Daniel Martinec, and Tomáš Pajdla. A convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4):407–422, August 2005.

[243] Tomáš Svoboda and Tomáš Pajdla. Epipolar geometry for central catadioptric cameras. *International Journal of Computer Vision*, 49(1):23–37, August 2002.

[244] A. Tate. I-N-OVA and I-N-CA - representing plans and other synthesised artifacts as a set of constraints, 2000.

[245] T Thueer, A Krebs, P Lamon, and R Siegwart. Performance comparison of rough-terrain robots - simulation and hardware. *Journal of Field Robotics*, 2007.

[246] T Thueer, P Lamon, A Krebs, and R Siegwart. Crab: Exploration rover with advanced obstacle negotiation capabilities. In *Proc. of The 9th ESA Workshop on Advanced Space Technologies for Robotics (ASTRA)*, 2006.

[247] Christian Thurau and Vaclav Hlavac. Pose primitive based human action recognition in videos or still images. In *CVPR 2008: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 8, Madison, USA, June 2008. IEEE Computer Society, Omnipress.

[248] M. Tomasello. Developmental perspectives on social intelligence and communication in great. In S.T. Parker and K.R. Gibson, editors, *Language and intelligence in Monkeys and Apes*, pages 247–273. Cambridge University Press, 1990.

[249] N Tomatis, S Bouabdallah, R Piguet, G Terrien, and R Siegwart. Autonomous navigation and security: A 13'000h/3'000km case study. In *Proc. of The 35th International Symposium on Robotics (ISR)*, 2004.

[250] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schultz. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, 35(4):460–470, 2005.

[251] D.R. Traum. *A Computational Theory of Grounding in Natural Language Conversation.* PhD thesis, Computer Science Department, University of Rochester, December 1994.

[252] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[253] S. Treue and J. C. M. Trujillo. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(575-579), 1999.

[254] R. Triebel, F. Dellaert, and W. Burgard. Using hierarchical EM to extract planes from 3d range scans. In *Proc. of the International Conference on Robotics and Automation(ICRA)*, 2005.

[255] R. Triebel, P. Pfaff, and W. Burgard. Multi-level surface maps for outdoor terrain mapping and loop closing. In *Proc. of the International Conference on Intelligent Robots and Systems (IROS)*, 2006.

[256] K.P Truong, D.A. van Leeuwen, and M.A. Neerincx. Unobtrusive multimodal emotion detection in adaptive interfaces: Speech and facial expressions. In D.D. Schmorrow and L.M. Reeves, editors, *Foundations of Augmented Cognition*, pages 354–363. LNAI 4565 proceedings, 2007.

[257] J.K. Tsotsos, S. Culhane, W. Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artifical Intelligence*, 78:507 – 547, 1995.

[258] E. Tunstel. Evolution of autonomous self-righting behaviors for articulated nanrovers. In *Proceedings of Fifth International Symposium on Artificial Intelligence, Robotics and Automation in Space (ESA SP-440)*, pages 341–346, Noordwijk, 1999.

[259] H. Uszkoreit, F. Xu, W. Liu, J. Steffen, I. Aslan, J. Liu, C. Müller, B. Holtkamp, and M. Wojciechowski. A successful field test of a mobile and multilingual information service system compass2008. In *Proceedings of HCI International 2007, 12th International Conference on Human-Computer Interaction*, Beijing, China, 2007.

[260] H. Uszkoreit, F. Xu, J. Steffen, and I. Aslan. The pragmatic combination of different cross-lingual resources for multilingual information services. In *Proceedings of LREC 2006*, Genova, Italy, May 2006.

[261] L. Van Breda, C. Jansen, and J.A. Veltman. Supervising uavs: Improving operator performance by optimizing the human factor. In *Proceedings International Conference on Human Computer Interaction - Augmented Cognition International*, 2005.

[262] S. Vasudevan, S. Gachter, and R. Siegwart. Cognitive spatial representations for mobile robots - perspectives from a user study. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) 2007 Workshop on Semantic Information in Robotics (SIR 2007)*, 2007.

[263] S Vasudevan and R Siegwart. Bayesian space conceptualization and place classification for semantic maps in mobile robotics. *Robotics and Autonomous Systems*, 2008.

[264] Shrihari Vasudevan, Stefan Gaechter, Viet Nguyen, and Roland Siegwart. Cognitive maps for mobile robots: an object based approach. *Robotics and Autonomous Systems*, 55(5):359–371, 2007.

[265] R. Volpe, J. Balaram, T. Ohm, and R. Ivlev. Rocky 7: A next generation mars rover prototype. *Journal of Advanced Robotics*, 11(4), 1997.

[266] M. Walker, D. Litman, C. Kamm, and A. Abella. Paradise: A framework for evaluating spoken dialogue agents. In *Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics (ACL'97)*, 1997.

[267] H. Wang, M. Lewis, P. Velagapudi, P. Scerri, and K. Sycara. How search and its subtasks scale in n robots. In *Proceedings of the 4th International Conference on Human Robot Interaction (HRI)*, 2009.

[268] M. White. Efficient realization of coordinate structures in combinatory categorial grammar. *Research on Language and Computation*, 4(1):39–75, 2006.

[269] B. Williams, M. Ingham, S. Chung, P. Elliott, M. Hofbaur, and G. Sullivan. Model-based programming of fault-aware systems. *AI Magazine*, 4(24):61–76, Winter 2004.

[270] Oliver Williams. Sparse bayesian learning for efficient visual tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1292–1304, 2005. Member-Andrew Blake and Member-Roberto Cipolla.

[271] S. Wilske and G.J.M. Kruijff. Service robots dealing with indirect speech acts. In *Proceedings of IROS 2006*, Beijing, China, 2006.

[272] N. Winters and J. Santos-Victor. Visual attention-based robot navigation using information sampling. In *Proceedings of IROS-2001*, pages 1670–1675, 2001.

[273] S. Wischmann and Pasemann F. From passive to active dynamic 3d bipedal walking - an evolutionary approach -. In M. Armada and P. Gonzalez de Santos, editors, *Proc. of the 7th Int. Conference on Climbing and Walking Robots (CLAWAR 2004)*, pages 737–744. Springer Verlag, 2004.

[274] J. M. Wolfe. Visual search in continuous, naturalistic stimuli. *Vision Research*, 34:1187–1195, 1994.

[275] J. M. Wolfe. Visual search in continuous, naturalistic stimuli. *Vision Research*, 34:1187–1195, 1994.

[276] Bo Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 1:90–97 Vol. 1, 17-21 Oct. 2005.

[277] H. Zender and G.J.M. Kruijff. Multi-layered conceptual spatial mapping for autonomous mobile robots. In Holger Schultheis, Thomas Barkowsky, Benjamin Kuipers, and Bernhard Hommel, editors, *Control Mechanisms for Spatial Knowledge Processing in Cognitive / Intelligent Systems*, Papers from the AAAI Spring Symposium, pages 62–66, Menlo Park, CA, USA, March 2007. AAAI, AAAI Press.

[278] Hendrik Zender, Geert-Jan M. Kruijff, and Ivana Kruijff-Korbayová. A situated context model for resolution and generation of referring expressions. In *12th European Workshop on Natural Language Generation (ENLG 2009)*, Athens, Greece, March 2009.

[279] Hendrik Zender, Óscar Martínez Mozos, Patric Jensfelt, Geert-Jan M. Kruijff, and Wolfram Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502, June 2008.

[280] L. Zettlemoyer and M. Collins. Online learning of relaxed CCG grammars for parsing to logical form. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 678–687, 2007.

[281] Karel Zimmermann, Jiří Matas, and Tomáš Svoboda. Tracking by an optimal sequence of linear predictors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):677–692, April 2009.