By Tamim Asfour, Mirko Wächter, Lukas Kaul, Samuel Rader, Pascal Weiner, Simon Ottenhaus, Raphael Grimm, You Zhou, Markus Grotz, and Fabian Paus

©ISTOCKPHOTO.COM/YAKOBCHUKOLENA

ARMAR-6

A High-Performance Humanoid for Human–Robot Collaboration in Real-World Scenarios

A major goal of humanoid robotics is to enable safe and reliable human–robot collaboration in real-world scenarios. In this article, we present ARMAR-6, a new high-performance humanoid robot for various tasks, including but not limited to grasping, mobile manipulation, integrated perception, bimanual collaboration, compliant-motion execution, and natural language understanding. We describe how the requirements arising from these tasks influenced our major design decisions, resulting in vertical integration during the joint hardware and software development phases. In particular, the entire hardware—including its structure, sensor-actuator units, and low-level controllers—as well as its perception, grasping and manipulation skills, task coordination, and the entire software architecture were all developed by one team of engineers. Component interaction is facilitated by our software framework ArmarX, which

---

Digital Object Identifier 10.1109/MRA.2019.2941246
Date of current version: 28 October 2019

1070-9932/19©2019IEEE

further facilitates the seamless integration and interchange of third-party contributions. To showcase the robot's capabilities, we present its performance in a challenging industrial maintenance scenario that requires human–robot collaboration, where the robot autonomously recognizes the human's need of help and offers said help in a proactive way.

Background and Related Work

Humanoid robots have existed for decades in research labs all around the world. Although there is a wealth of possible applications in practical use cases, the inherent complexity of humanoid robots as well the intricate nature of the tasks they are expected to perform has hindered their transition from research labs to real-world use. One of these difficult tasks is the need to work closely with humans in a safe, predictable, intuitive, and productive manner. An important step toward the transition of humanoids into real-world applications is therefore the proliferation of technologies that permit robots to work cooperatively and safely with humans, which has led to the development of collaborative robots.

On the design side, these technologies include lightweight robot design [1]; anthropomorphic kinematics (e.g., [2] and [3]); force and torque sensing [4]; high-speed, real-time bus systems [e.g., Ethernet for control automation technology (EtherCAT)]; and an extremely high level of system integration, which is required to fit the functionalities into the desired form factor.

These technologies have found their way into commercial products in the form of individual arms (e.g., by Kuka, Universal Robots, and Franka Emika), dual-arm systems (e.g., by ABB and Kawada Industries), and even complete humanoid robots such as PAL Robotics' REEM-C [39] and TALOS [5].

The design, however, is only one aspect of truly collaborative robots. Other crucial aspects are the cognitive abilities of the robot, which are needed to correctly interpret situations and understand how to work together with humans, even in environments that are not completely known beforehand. A hardware platform that supports cognitive, bimanual mobile manipulation capabilities must be equipped with various exteroceptive sensors and sufficient on-board computing resources, in addition to the body, arms, and end effectors. Mobility is another important requirement for such a robot because it drastically expands the working radius, thereby enabling a vast number of additional applications.

Several complete humanoid systems that integrate these capabilities for real-world use have recently been introduced. The list of such systems is long; therefore, we limit the scope of this article to systems designed for dual-arm mobile manipulation. These include the HRP (Humanoid Robotics Project) robots, used for aircraft assembly tasks [6] and construction work [7]; DLR's Rollin' Justin [8]; the fully torque-controlled TORO [9]; and robots such as the WALK-MAN [10] and E2-DR [11].

With ARMAR-6, we set out to advance the state of the art in mobile collaborative robots in terms of its system design, i.e., regarding its physical capabilities and overall technological readiness for real-world applications, as well as in all aspects of the cognitive capabilities necessary to accomplish challenging tasks beyond laboratory environments. The presented robot advances the state of the art with respect to kinematics, workspace, and workload. To the best of our knowledge, no other humanoid system currently exists that combines an arm reach of more than 1 m with a carrying capacity of more than 10 kg at full arm extension to go along with its limitless joints. ARMAR-6 is not only physically capable of demanding tasks; it is also equipped with a comprehensive suite of sensors as well as the cognitive abilities necessary to facilitate natural, safe collaboration with humans.

Depicted in Figure 1 is ARMAR-6, the latest generation in the ARMAR humanoid robot family [15]. It was developed as a robotic research platform within the SecondHands project [40], the vision of which is to develop and enable a robot to literally provide a human worker with a second pair of hands to support, increase safety, and enhance efficiency. Its development is driven by an ambitious use case, in which the robot autonomously and proactively assists a maintenance technician with repair tasks on a material handling system in an online retailer's highly automated customer fulfillment center. This particular scenario poses many requirements of the robotic system, which challenged us to "push the envelope" of what is currently possible in collaborative robotics.

Its deployment in an actual warehouse environment requires ARMAR-6 to be extraordinarily robust and reliable. The robot therefore incorporates all of the experiences from the construction and deployment of its predecessors regarding the choice of components, material selection, cable routing, and software modules on all levels to realize a

Figure 1. ARMAR-6: a collaborative humanoid robot.

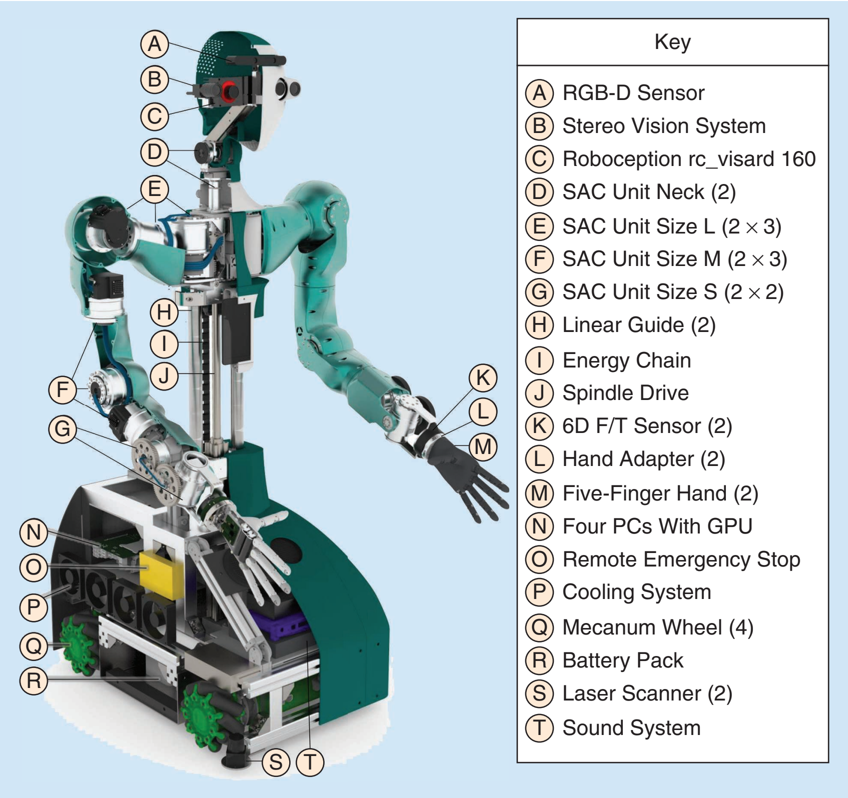


Figure 2. A visualization of essential internal components (i.e., Ⓐ–Ⓣ) of ARMAR-6. RGB-D: red-green-blue-depth; SAC: sensor-actuator-controller; F/T: force-torque.

highly reliable system. In addition, we designed the robot, in terms of both its hardware and software, to be highly modular. Key components can be exchanged quickly, and subsystems of the robot, such as its wheeled platform and its arms or head, can easily be tested and used separately. Because some of the tasks involve handling heavy machine parts, e.g., the overhead covers of conveyor belts, the robot requires a comparatively high payload capacity. Concretely, we designed ARMAR-6 to handle up to 10 kg in each hand. The tasks require the ability of overhead manipulation well above 2 m as well as picking up objects from the floor. Mobility while carrying heavy objects and the ability to handle tools made for humans are other key requirements.

In terms of cognitive abilities, ARMAR-6 must be able to understand a variety of scenes, recognize human actions and intents, learn from human observation and its own experiences, and process the data produced by cameras, laser scanners, and proprioceptive sensors while still acting fluently at human-level speeds. It must be able to apply large amounts of force when needed, while at the same time interacting gently and predictably with humans, e.g., in the context of kinesthetic teaching.

All of these requirements can only be fulfilled when addressed jointly in the mechanical design of the robot, its actuators, control systems, and computing hardware as well as its functional architecture and efficient implementation in the software, which distributes tasks, ensures reliability, and allows for adaptation to new tasks and scenarios with little additional programming.

The control software has access to the sensors and actuators of the arms, head, and mobile platform via the high-speed EtherCAT bus at a rate of 1 kHz, which enables the fast and convenient prototyping of novel control methods. Robot control on all levels is supported by well-documented software interfaces implemented in the ArmarX software framework [16]. Moreover, the robot's arms provide standardized interfaces, which comply with the International Standards Organization's 9409-1-50-7-M6 to quickly exchange end effectors (e.g., different hands or grippers), assuming they are 48-V compliant and have an EtherCAT interface.

We introduced the humanoid robot ARMAR-6 in [17]. In this article, we elaborate on the reasoning behind the design choices for developing a humanoid maintenance assistant for man-made environments and on the hardware and software codesign. We further put into perspective all of our separately published contributions with respect to the overall design, control, and software architecture of the ARMAR-6 robot.

Hardware and Software Codesign for Embodied Intelligence

The broad set of requirements imposed by the aforementioned challenging scenario necessitates a principled approach to the design of the robot's hardware and its control software. We chose a fundamentally modular design approach for both the hardware and software of ARMAR-6 to enable separating the design into manageable parts. On the hardware side, this modular approach means that each part of the robot can be operated separately and maintains a well-specified interface that integrates with the other components. On the software side, components can be distributed over different machines to ensure the high performance and responsiveness needed for collaborative tasks.

Overall Robot Design

ARMAR-6 has 28 active degrees of freedom (DoF), a maximum height of 192 cm, and a weight of 160 kg (plus a 40-kg battery pack). On its highest level, the robot's hardware is split into five different modular robotic systems, including the dual-arm system, five-finger hands, sensor head, torso, and mobile platform. An overview of these components and their internal workings is displayed in Figure 2. The following sections elaborate on the development of these parts as well as the software that controls them. A summary of the hardware is listed in Table 1.

Table 1. Specifications and key hardware components of ARMAR-6.

Weight	–	–	–	160 kg (without batteries)
Height	–	–	–	152–192 cm
Footprint	–	–	–	60 × 80 cm
Platform speed	–	–	–	1 m/s
Payload	–	–	–	10 kg per arm
Power consumption	–	–	–	460 W (nominal), 1,050 W (peak)
DoF	Neck	2	Custom SAC units based on brushed dc motors, harmonic drive-reduction gears, and custom motor controllers	Pitch and yaw
–	Arms	2 × 8	Custom SAC units [18] based on Robodrive BLDC motors, harmonic drive-reduction gears, and ELMO motor controllers	Four per shoulder, two per elbow, and two per wrist
–	Hands	2 × 14	Maxon brushed dc motors	Underactuated five-finger hands with one motor for 12 finger joints and one for two thumb joints per hand
–	Torso	1	Dunker BLDC motor with a spindle drive, brake, and ELMO motor controller	Linear actuator with 40-cm vertical travel
–	Platform	4	Donkey-motion drive system with BLDC motors, a brake, and an ELMO motor controller.	Omni wheels for holonomic motion
Sensors	Head	1	Roboception rc_visard 160 stereo sensor	3D vision
–	–	1	PrimeSense Carmine 1.09	RGB-D
–	–	2	Point Grey Flea3 USB-3 cameras	Stereo vision
–	SAC units	1	Renishaw AksIM MBA7/8	Absolute position
–	–	1	AMS5306	Incremental position
–	–	1	Texas Instruments ADS1220 analog-to-digital converter	Torque (strain gauge full bridge)
–	–	1	Bosch BNO055 IMU	Acceleration and rotational rate
–	–	5	Various temperature sensors	Temperature of different components
–	Arms	2	ATI mini 45 F/T sensor	6D F/T sensing
–	Torso	1	Waycon draw wire sensor SX50	Absolute vertical position
–	Platform	2	Hokuyo UST-10LX 2D lidars	Laser range finders for orientation and safety
Power supply	Internal	–	Nickel metal hydride battery with 38 AH at 48 V (1,824 Wh)	Power autonomy for up to 4 h
–	External	–	48-V power supply	Tethered power for development and testing
Computers (4)	Real time	–	Mini-ITX, Core-i7, 32-GB RAM, Ubuntu, and ArmarX	RT-PREEMPT patch and Simple Open EtherCAT Master (SOEM)
–	Vision	–	–	GeForce GTX-1080 GPU
–	Speech	–	–	Roland USB sound card
–	Planning	–	–	ArmarX master node
Communication	Internal	–	EtherCAT Bus	Real-time automation bus connecting all actuators to the real-time computer; a 1-kHz bus update rate
–	External	–	Gigabit Ethernet	Either via LAN, 2.4 GHz, or 5-GHz WLAN via the internal router and switch
Robot-development environment	–	–	ArmarX	Middleware, comprehensive high- and low-level APIs
User interface	–	–	ArmarX	GUI, natural language, and various programming languages

RAM: random-access memory; IMU: inertial measurement unit; LAN: local area network; WLAN: wireless LAN; APIs: application programming interfaces; GUI: graphical user interface; BLDC: brushless dc.

Dual-Arm System With 16 Torque-Controlled DoF

Inspired by human kinematics, each arm has 8 DoF, including a clavicle joint of the inner shoulder [19]. This additional joint significantly increases the bimanual workspace of the dual-arm system [20]. In combination with a reach of 1.3 m and a large range of motion in each joint, the dual-arm system has a total workspace of 8.4 m^3 and a bimanual workspace of 4.9 m^3. The structure of the arms follows an exoskeleton design approach: hollow aluminum parts connect the joints mechanically and simultaneously serve as a support structure and covers. This results in a stiff yet lightweight construction, enabling the robot to carry loads up to 20 kg at full arm extension.

To generate the required torque, each arm joint is actuated by a highly integrated sensor-actuator-controller (SAC) unit [18]. These modular SAC units (Figure 3) include a brushless dc motor and a backlash-free Harmonic Drive reduction gearbox with all the necessary bearings as well as a comprehensive sensor suite, communication, and motor control electronics. The incremental and absolute position encoders with an accuracy of 0.1° and a resolution of 19 b allow for highly precise position and velocity control. To aid safe human–robot interaction (HRI), we pursued an active compliance approach; therefore, each SAC unit includes a strain-gauge-based torque sensor with a sensitivity of more than 0.04 Nm. Combined with fast, real-time sensor signal transmission provided by the EtherCAT bus, this facilitates precise torque control at a rate of 1 kHz.

In highly integrated mechatronic systems, cabling is crucial for overall system reliability. We used slip rings, which sanction continuous cable rotation between rotating and nonrotating parts of the joints. As a result, each arm has three limitless joints, i.e., the rotation of the shoulder, upper arm, and forearm.

In addition to high integration and robustness, modularity and encapsulation are key design principles for SAC units. Electrically, the daisy-chained SAC units expose minimal interfaces: The input consists of connectors for the communication and dc power bus as well as a separate emergency stop (e-stop) cable. The output comprises the corresponding three connections. In between the units, the three cables (i.e., communication, power, and e-stop) are routed inside the hollow structure, thus protecting them from mechanical stress and external interferences. The high modularity also simplifies the maintenance of the robot arm because the SAC units can easily be exchanged if needed. To achieve or even surpass human performance with anthropomorphic appearance and kinematics, the units were designed to be scalable: the three different SAC unit types afford peak torques between 63 and 176 Nm at maximum rotational speeds between 79 and 206°/s.

The final SAC unit in the wrist is followed by an additional 6D force-torque (F/T) sensor, which provides the robot with haptic feedback during manipulation and HRI tasks. It is attached to a Schunk hand adapter that offers a well-defined mechanical and electronic interface. This allows for the quick

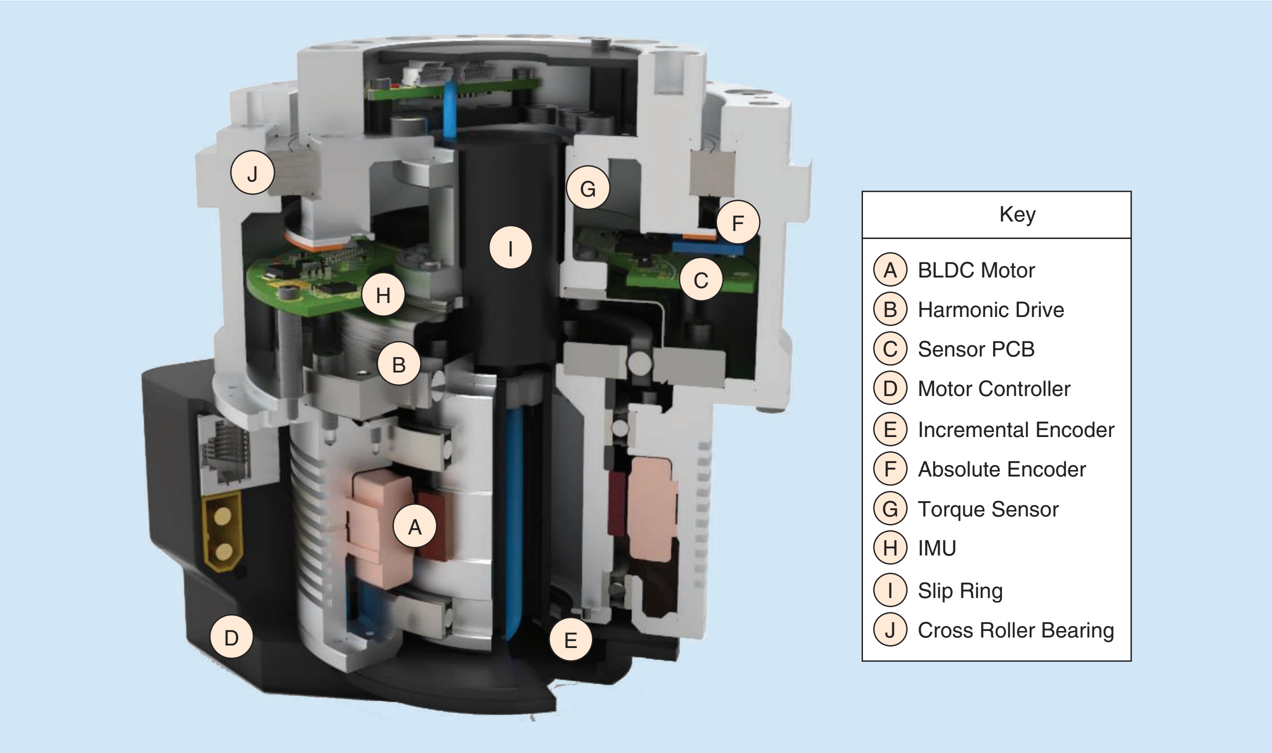


Figure 3. A rendered 3D cross section of the midsized SAC unit. The dual-arm system of the ARMAR-6 has 16 highly integrated SAC units in three different sizes. PCB: printed circuit board.

exchange of different end effectors and adapts the robot to new scenarios.

Underactuated Five-Finger Hands

For operating in human-centered environments, ARMAR-6 is equipped with two five-fingered hands. Its anthropomorphic shape and kinematics based on human proportions permit the manipulation of different tools and other man-made objects. Each of the 14 finger joints (two in the thumb and three per finger) is supported by two ball bearings. Actuation of the joints is realized using Dyneema tendons guided through polytetrafluoroethylene tubes for reduced friction.

The four fingers and the thumb are actuated by one motor. The motor torque is distributed to the fingers using an adaptive underactuation mechanism based on the TUAT/Karlsruhe mechanism [21]. This mechanism helps the fingers wrap around objects and adapt automatically to their shapes.

Because the hands are used in maintenance tasks involving heavy objects, they must be simultaneously robust and light to reduce the load on the arm. Hence, most of the structural hand parts are 3D printed from durable polyamide using selective laser sintering. All of the cables are hidden in the interior of the hand to avoid electric failures. The hand itself is covered by a protective work glove.

Sensor Head

The sensor head equips the robot with a range of visual sensors and includes five cameras, divided into three independent systems. The first is a Roboception rc_visard 160, a 16-cm baseline stereo camera system with on-board depth image processing. It is used for peripheral vision with an optimal depth measurement range from 0.5 to 3 m. The second stereo camera system, a pair of Point Grey Flea3 cameras, has a baseline of 27 cm and is used for foveal vision. These two cameras are placed on both sides of the Roboception sensor with parallel image planes. The third system is a PrimeSense active red-green-blue-depth (RGB-D) sensor, which is particularly advantageous for unicolored, featureless surfaces.

The camera systems are mounted on a common aluminum frame, whose direction is controlled by a pan-tilt unit in the neck. Its two joints are driven by custom-built, modular sensor-actuator units, a new, miniaturized version of SAC units. Hence, the modules combine the same highly precise absolute position encoder with a backlash-free Harmonic Drive reduction gearbox and stiff structure to achieve high positioning accuracy of the camera systems. In combination with its speed of up to 510°/s and its high control frequency aided by EtherCAT, the sensor head is precise and fast, enabling effective pose tracking and gaze stabilization. Similar to the hands, the head and neck are protected by 3D printed cover parts.

Mobile Platform and Torso

To assist the robot in supporting technicians at different areas of the warehouse, mobility is a key requirement. We designed an omnidirectional, wheeled mobile platform that meets all the requirements for an industrial workplace. The platform offers a space for all of the necessary components, allowing the robot to operate autonomously and without external cables. It contains four computers and a GPU for on-board computing, a sound system, network peripherals, the power management system, and a 1.8-kWh battery pack that permits up to 4 h of power-autonomous operation under nominal operating conditions. All of the components are covered and protected by a fiberglass housing.

To increase the robot's workspace, it has an extensible torso realized as a single prismatic joint, which is driven by a dc motor with a spindle drive and guided by two parallel linear bearings. The torso joint permits the robot to change its height by 40 cm, which increases the workspace of the dual-arm system from 8.4 to 10.7 m^3 (Figure 4). As a result, ARMAR-6 is able to pick up tools from the ground and can hand a tool to a technician on a ladder at a height of 2.4 m. A brake keeps the height of the shoulders constant when necessary, thereby ensuring energy-efficient actuation. The absolute height of the torso is determined by a draw wire sensor. All of the cables between the upper body and platform are routed through an energy chain that protects them and avoids mechanical stress.

For safe HRI, in addition to its comprehensive sensor setup of the dual-arm system (in particular, the torque and F/T sensors), the brakes in its torso, and its mobile platform, ARMAR-6 also features a wireless and a wired remote e-stop. The platform further contains two planar laser scanners that provide 360° distance information, which is used for safety, obstacle detection, and map-based navigation.

Functional Software Architecture

The software architecture of ARMAR-6 satisfies similar requirements as the hardware, i.e., modularity, reusability, and

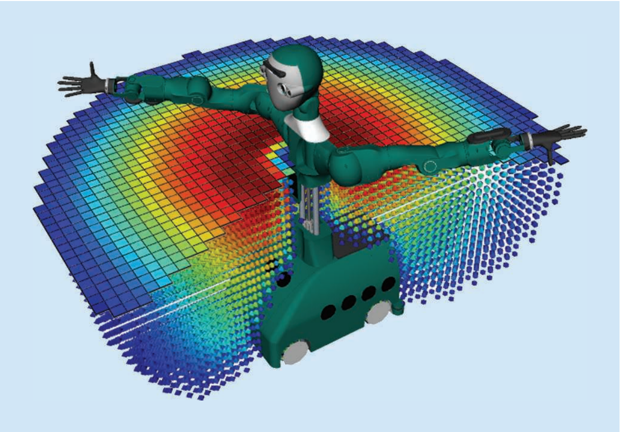


Figure 4. A sectional view through the 10.7 m^3, nearly ellipsoidal workspace of the dual-arm system that incorporates the prismatic joint of the torso. The color of the voxels corresponds to the local manipulability as described by Yoshikawa [22], with an extension so that the manipulability maps for both arms are merged into one manipulability map. The warmer the color, the higher the manipulability.

scalability. All of our recent humanoid robots [2], [23] share the same cognitive and functional software architecture implemented in ArmarX [24], [41]. As a complex humanoid robot system, ArmarX is not only a robot middleware but also a complete functional cognitive architecture offering ready-to-use modules on multiple levels of abstraction. Figure 5 shows the key modules of the architecture, which are divided into three levels of abstraction: 1) real-time control and hardware abstraction, 2) perception and the robot memory system, and 3) task planning.

On the lowest level, nearly all data are subsymbolic. The hardware abstraction layer connects with the robot-agnostic, real-time control framework. This abstraction level provides its data to the upper levels and receives commands from the highest level. On the middle level, perception modules for sensor processing such as object localization, self-localization, and human pose estimation are located. Their results are stored in the ArmarX memory system, MemoryX, which includes modules such as the working memory for a consistent world state, long-term memory for self-experienced data, and prior knowledge that contains data provided by robot programmers (e.g., 3D object and world models) as well as preprogrammed trajectories. The result of the sensor processing is usually represented in (but not restricted to) easy-to-process formats such as 6D poses or symbolic relations. On the highest abstraction level, reasoning modules, such as natural language understanding, symbolic planning, and statecharts, coordinate the robot subsystems primarily on a symbolic level.

Figure 6 displays an abstract depiction of an exemplary statechart (in this case, for the handover task described in the "Human Activity Recognition and Active Handover" section) and the interconnections between respective perception and the control modules. All of the modules run in parallel and provide their data asynchronously to the statecharts, which generate targets for the controllers that run synchronously in the real-time layer.

One key principle of the robot architecture is the modularity of the software based on common interfaces. This facilitates the easy exchange or insertion of new software modules to be used on multiple robot platforms, which is especially important for the purpose of research. Due to this modularity, failures of individual modules do not affect the complete system. The complexity of the robot's software system results in a high number of interconnected modules that create a complicated network of dependencies. To provide an assessment of the system's state, ArmarX presents

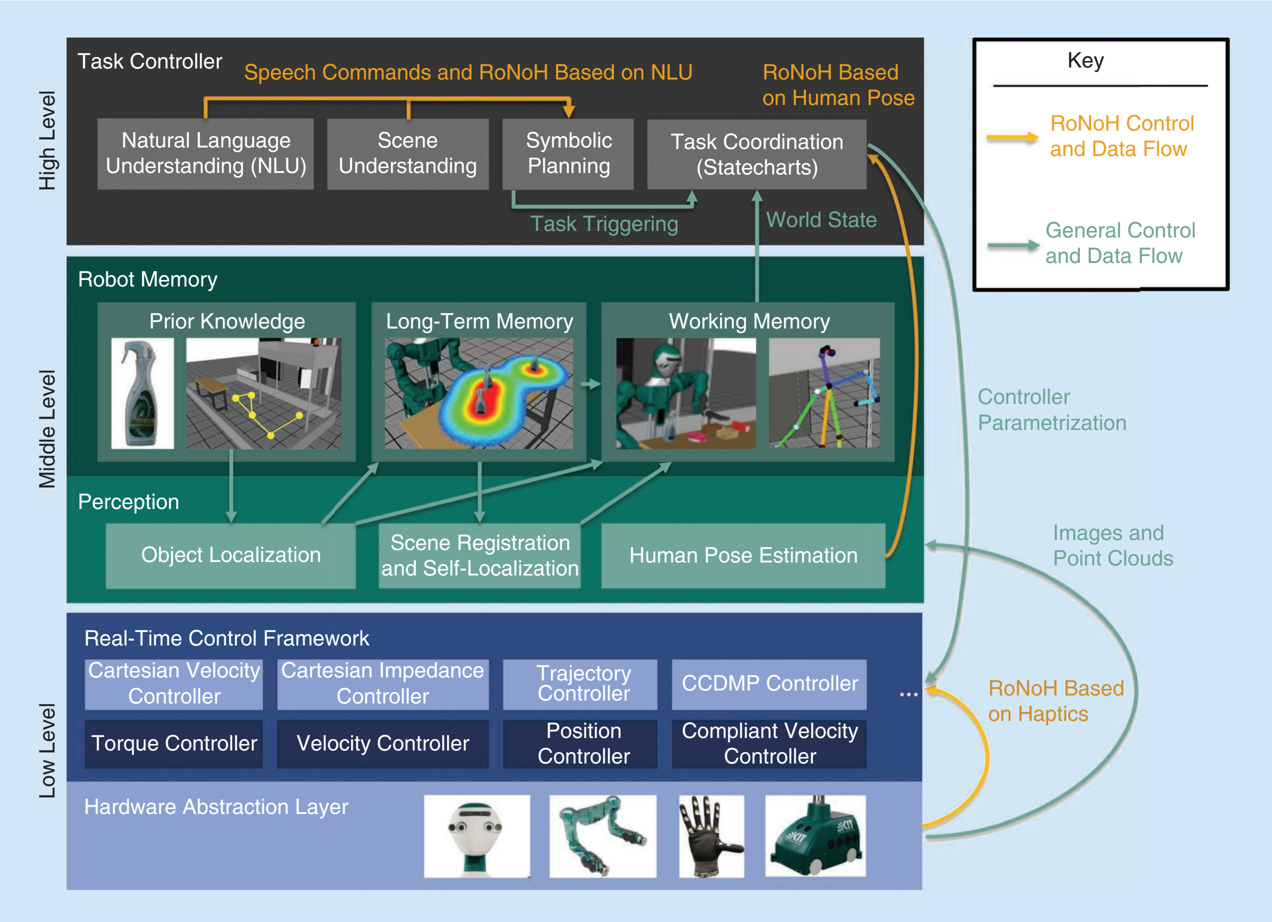


Figure 5. An overview of ArmarX, the functional software architecture of ARMAR-6. The image is divided into its three main conceptual levels and their principal interconnections. The information is increasingly abstracted from low (e.g., raw sensor values) to high levels (e.g., symbolic predicates). RoNoH: recognition of the need of help; CCDMP: Coordinate Change Dynamic Movement Primitive.

various monitoring tools at the sensor and module levels. This enables advanced safety features: if any software module configured as "system-relevant" stops working, all robot actuators go immediately to an e-stop, where they hold their current positions with reduced torque.

To allow the reuse of modules on different robot hardware platforms, hardware abstraction is realized on two levels. The first is a generic, high-performance real-time control framework in ArmarX, which affords implementations of general real-time motion controllers. These can be used specifically for different robots in a synchronized manner. They are accessible to other modules via the middleware. On the second level, simple and unified interfaces offer basic control functionality and make all sensor values available for modules in the robot network.

Although the real-time control framework is designed for a general purpose, it was codesigned with the specific hardware of ARMAR-6 to ensure that the framework overcomes the constraints of the real hardware. Because ARMAR-6 uses the high-performance real-time bus EtherCAT, the software framework must be able to fulfill real-time constraints while still being extensible and modular. This is mainly achieved by template programming and inheritance throughout the whole framework

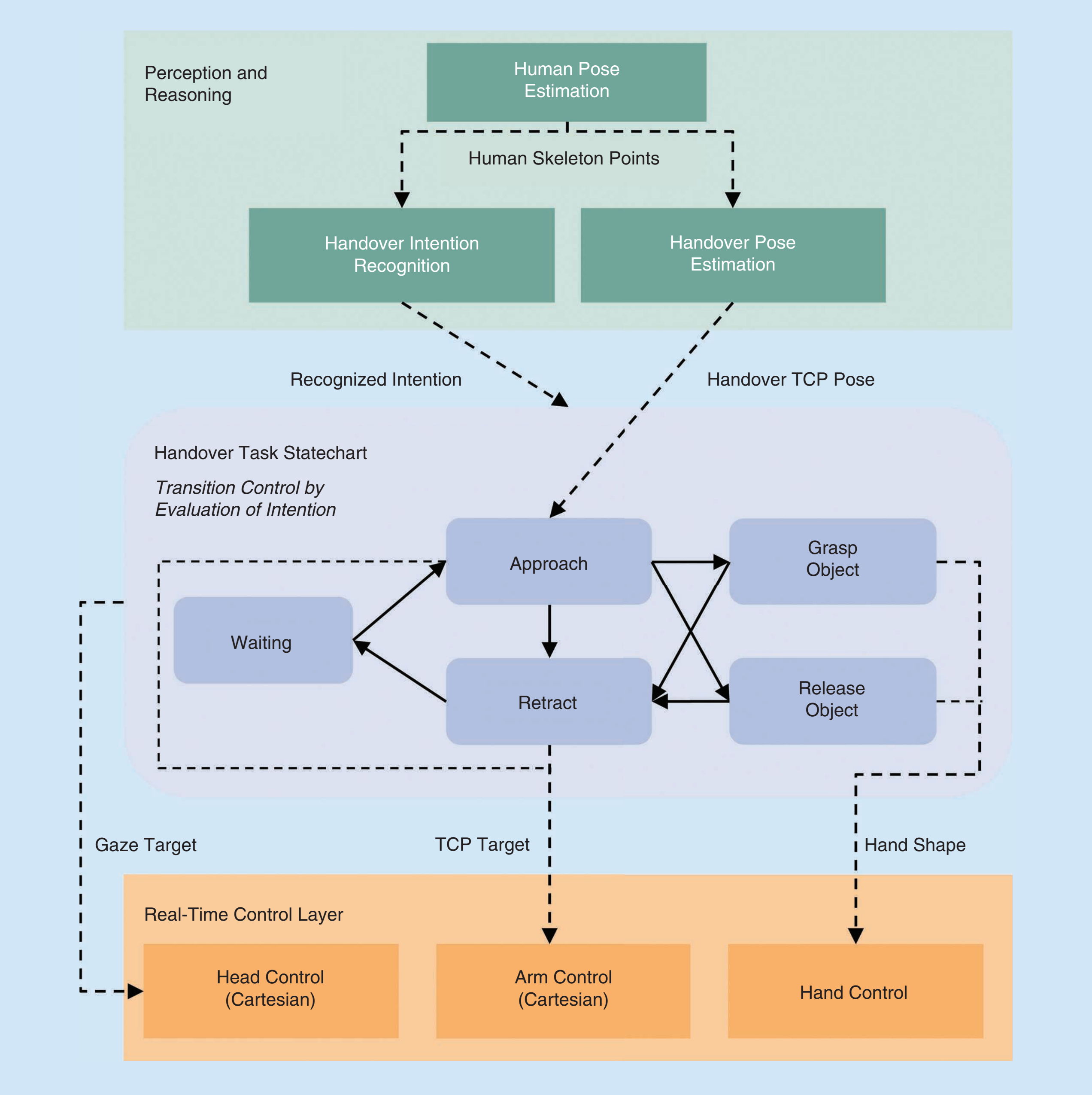


Figure 6. A flowchart of the handover task, showing the coordinating statechart as well as the involved perception and control modules needed for an interactive and adaptive handover. TCP: Transmission Control Protocol.

as well as by preallocating all resources needed for the current hardware setup.

With this real-time control framework, ARMAR-6 offers state-free, basic controllers not only for individual joints such as position, velocity, or torque controllers but also for multijoint Cartesian controllers as well as controllers for trajectory execution, impedance control, and dynamic movement primitives (DMPs). Furthermore, the control framework provides unrestricted access to all sensor values and low-level controllers in a synchronous control loop at 1 kHz. More complex or non-real-time controllers run their own control loop in a separate thread and synchronize their control targets in a lock-free manner with the real-time control loop.

For scalability, ArmarX relies on a multithreaded, distributed design concept of modules. Every module can run transparently on any machine in the robot network. Additionally, there is no central processing loop, which permits it to run every module at individual frequencies. This ensures that the large number of components (i.e., the sensors and actuators) of ARMAR-6 does not affect the robot's overall performance.

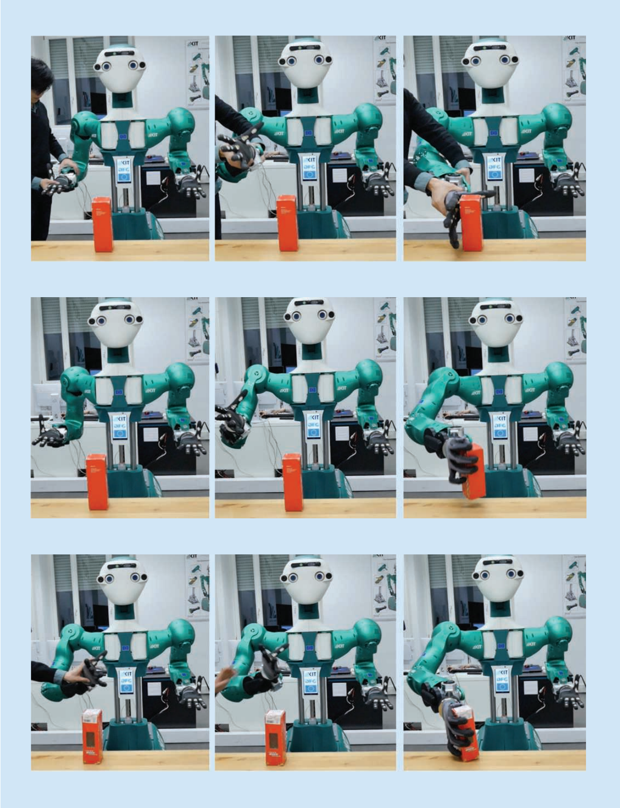
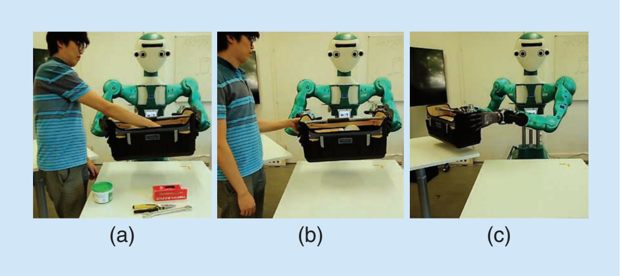


Figure 7. Kinesthetic teaching and compliant-motion reproduction based on task-space impedance control. During motion execution, the robot is able to compliantly adapt its actions to external-force perturbations and safely interact with humans.



(a) (b) (c)

Figure 8. The compliant bimanual control system allows for the manipulation of large and heavy objects. (a) The robot holding a toolbox filled with objects totaling 6 kg. (b) The coupling of both arms is ensured by the bimanual controller during perturbations. (c) The motion is encoded using CC-DMP, and the forces exerted on the toolbox are generated by the bimanual controller.

Autonomous Abilities for Physical Human–Robot Collaboration

The hardware and software components described in the previous section endow ARMAR-6 with the abilities required in the context of safe physical human–robot collaboration. This section will highlight a few of them in isolation, before we describe an integrated validation scenario leveraging a variety of these abilities. All of the abilities are implemented as software modules, statecharts, or real-time controllers within ArmarX.

Task-Space Impedance Control and Bimanual Control

Compliant control is a crucial part of a collaborative robot's capabilities that allows for safe HRI. We developed a task-space impedance-control scheme using joint-level torque control that tolerates compliant arm motions [25]. We modeled the robot's hand motion through the use of a damped spring system guided by an attractor to a moving target. These targets are generated by a DMP learned in either joint or task space. Figure 7 depicts the task-space impedance-control procedure during a grasping motion, which was learned from kinesthetic teaching. In particular, we investigated two methods for carrying large and heavy objects (bimanual manipulation tasks). The first method, Coordinate Change DMP (CC-DMP) [26], combines two DMPs that describe the motion of each arm coupled in a leader-follower manner. The second method is based on force-admittance control, as described in [27]. An example task with the bimanual controller is shown in Figure 8, where the robot holds a large, heavy object and puts it to the side. The motion is encoded by CC-DMP, while the bimanual controller guarantees persistent coupling, even during external perturbations.

Grasping of Known and Unknown Objects

ARMAR-6 can grasp known, familiar, or unknown objects. For known objects, grasps are determined offline using a grasp planner, which exploits topological object information (represented as an object skeleton) to identify suitable grasping regions. It selects the grasp type and aligns the hand according to the local skeleton and surface information [28]. Subsequent motion planning includes the planning of time-optimal, collision-free trajectories for fast grasping as well as a human-like approach direction and hand orientation.

For grasping familiar objects, part-based grasping methods are used to generate grasps based on the object shape by employing machine-learning methods, which are trained on a set of various object classes [29], [30]. In the case of unknown objects, ARMAR-6 combines a deep-learning grasping approach with reactive-grasp execution [31]. Based on fused tactile and visual data, a neural network generates several grasp candidates with associated grasp success probability ratings. To cope with model uncertainty while executing the best-rated grasp candidate, a reactive-grasping scheme is applied. To this end, the 6D F/T sensor in its wrist is used to infer the contact point between its fingers and the object. Based on the force sensor data, its grasp pose is iteratively refined, adjusted, and executed.

Human Activity Recognition and Active Handover

One central aspect of HRI is to recognize and predict human actions. To estimate the 3D human pose, we employ state-of-the-art, marker-free methods [32], [33], which use RGB-D images as input and run on the on-board GPU at ~15 h. A neural network classifier utilizes the human pose to estimate the human's current activity. Additionally, the human pose provides input for the active handover skill of the robot, which is implemented using the handover controller described in [34]. This controller receives the human hand pose to adapt the approach motion during the handover task. Force feedback from the 6D F/T sensor in the wrist is used to trigger hand opening and closing actions in robot–human and human–robot handover tasks. Figure 9 shows the 3D visualization of the working memory of the robot and an overlay skeleton on the input image during the handover execution. Figure 6 displays the software modules involved in the handover task.

Recognition of the Need of Help

One of the key aspects of a robot assistant is the ability to recognize that a human coworker needs help and deliver said help proactively.

In our work, we investigated how recognizing the need of help can be implemented based on different sensory modalities furnished by the robot, including visual information, haptic and force information, and speech and dialog. Using natural language affords an intuitive way for communicating the need of help either by employing speech commands to directly request help or based on the task-specific interpretation of utterances supplied by the natural language component. Vision-based human pose tracking and activity recognition deliver a second way of identifying the need of help during task execution. Given a description and all the required objects and actions of a collaborative task usually performed by two humans, the robot can deduce that a second agent is missing by interpreting the current scene. This allows the robot to proactively offer help by executing the task of the missing agent. Another way for detecting the need of help is the use of force and haptic information in collaborative human–robot manipulation tasks, such as jointly carrying a large object. Based on irregular force patterns and/or sudden changes in these patterns during the task, the robot is able to surmise that the human collaboration partner is struggling, and a regrasping action may furnish help as it will reduce the load on the human. We implemented and integrated the strategies discussed in this section and demonstrated their performance in several complex warehouse maintenance tasks.

Natural Language Understanding

The natural language dialog system of ARMAR-6 enables intuitive human–robot communication. The speech recognition and dialog system is a deep-learning architecture trained in an end-to-end fashion. It is implemented as an attention-based encoder-decoder network [35], while the natural language understanding is refined using a multitask learning approach [36]. ARMAR-6 can react to a large variety of direct commands (e.g., " Bring me the spray bottle") but can also infer instructions from spoken language utterances to suggest providing help in a task or ask for missing information. When the technician tells the robot that the conveyor drivetrain must be cleaned, e.g., ARMAR-6 will conclude that cleaning fluid is needed and will assist the technician by bringing it and handing it over.

Assisting Humans in Maintenance Tasks

To validate the capabilities of ARMAR-6 against a realistic benchmark, we devised a challenging demonstration scenario that requires advanced scene understanding, human intention recognition, mobile manipulation, and physical HRI.

Demonstration Scenario

Derived from the goals of the SecondHands project, the scenario represents the routine inspection and maintenance of an overhead conveyor system in an automated customer fulfillment center. This work is typically carried out by two human technicians, one of whom is the expert while the other assumes the role of a subworker. This latter role is physically demanding, and the cognitive tasks involved make it challenging for humanoid robots. In our scenario, ARMAR-6 is the subworker tasked with seamlessly and effectively

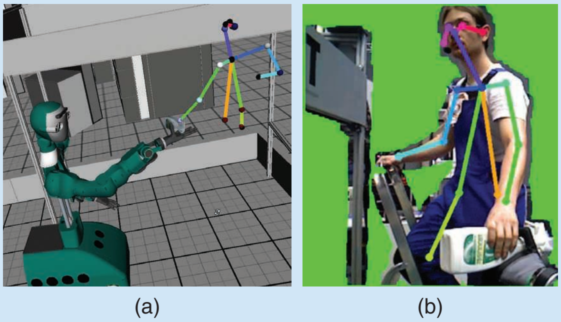

(a) (b)

Figure 9. (a) The active handover of a tool between ARMAR-6 and the human technician. (b) The robot uses marker-free 3D human pose estimation and force feedback to recognize the human's intention and so provide help when needed.

supporting the expert technician. A graphical overview of the scenario as well as the robot's tasks, actions, primary sensor modalities, and methods of HRI is represented in Figure 10. An unedited video of this scenario is available online [42]. The scenario is split into two main parts. These two parts, i.e., removing the cover panel and fetching a spray bottle, are very different in nature. Although the first part requires very close, seamless, and compliant physical human–robot collaboration, the second part requires mostly autonomous mobile manipulation abilities, including robust and comprehensive scene perception and interpretation.

Compliant Collaborative Manipulation

The technician realizes that the conveyor system is defective and, using plain English, informs ARMAR-6. By utilizing the continuously active automatic speech recognition (ASR) system [37], ARMAR-6 understands the technician's instructions and deduces the imminence of a maintenance task that it is able to assist with. This proactive, multimodal recognition of the need of help (RoNoH) is one of the key cognitive abilities of the SecondHands project that we are investigating, as it represents a crucial enabling technology for seamless HRI. Having realized that help is needed, the robot follows the technician, who proceeds without waiting to inspect the defective conveyor system.

When the technician starts to remove the cover panel of the conveyor to gain access to the drive train [depicted in Figure 10(a)], the robot recognizes this action and concludes that its help, in the form of supporting the cover panel on the opposite end, is needed. This recognition is based on visual 3D human pose tracking [32], [33] rather than natural language understanding. Using its laser-based self-localization system, the robot finds the right position under the conveyor and raises both hands to support the panel. ARMAR-6 relies on its tactile sensing (enabled by its wrist-mounted 6D F/T sensors) to gently make contact with the cover panel and grasp it. Because the geometry of the cover panel is known to the robot, suitable grasps for both hands and motion

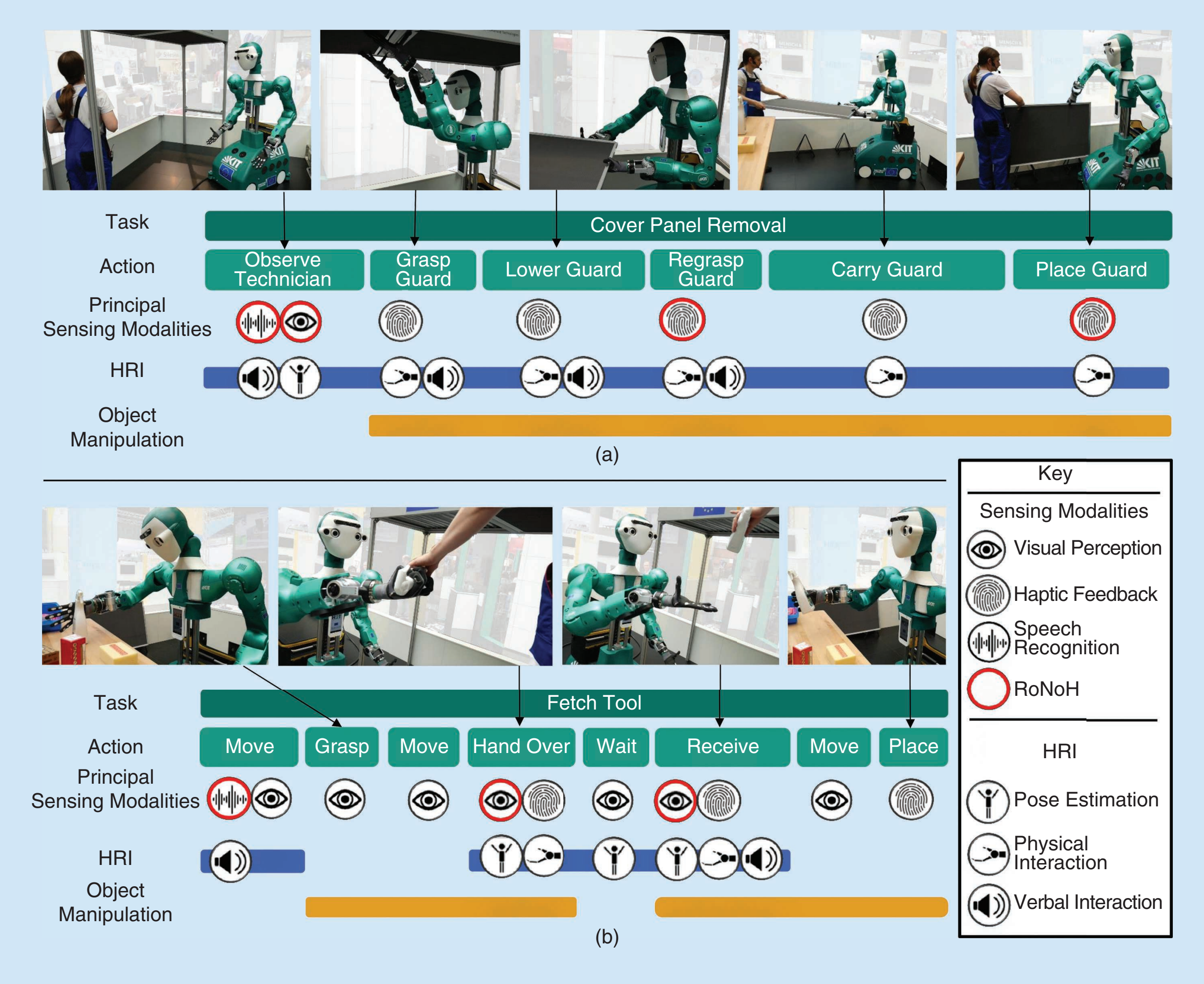


Figure 10. The warehouse demonstration scenario broken down into tasks, actions, principal sensor modalities, HRI modalities, and object manipulation. (a) The compliant collaborative manipulation is shown in a cover panel removal task. (b) Autonomous mobile manipulation is demonstrated by the robot's retrieval of a tool for the technician.

sequences for both arms are defined as prior knowledge. When it is in place, ARMAR-6 verbally confirms that it is ready to lower the panel. As soon as the technician has fully unmounted the panel, which the robot again realizes through haptic feedback, the human–robot team jointly lowers it. Once lowered, the robot confirms in natural language that it is ready for the next step. To carry the panel to a temporary location, the technician must step around the posts of the conveyor setup, which requires him to grasp the cover panel on one of its corners. ARMAR-6 realizes that the panel is out of balance and concludes through tactile sensing that it can help the technician by regrasping it on the opposite corner. It does so and again verbally confirms its readiness for the next step.

The technician then starts to walk toward the intended placing location of the cover panel. ARMAR-6 has no a priori notion of this location. Instead, it is guided by the movement of the technician, which it senses through the displacement of its own hands. This force-based method for guiding is enabled by joint-level compliance in every joint of the robot's arms, which is one of the principal joint control modes of ARMAR-6 and has proven to be invaluable for physical HRI.

Once the cover panel is at the intended location, the technician begins a placing movement. The robot senses the onset of this motion with its wrist-mounted 6D F/T sensors, recognizes that it can assist the technician with placing the heavy object, and, in turn, starts the placing motion. Because the robot's arms are still compliant, the technician can determine the exact placement position while the robot simply supports the weight of the panel.

Autonomous Mobile Manipulation

After the cover panel is removed, the technician inspects the exposed inner workings of the conveyor and decides that the drive train must be cleaned. Through the use of natural language, the technician expresses what needs to be done by saying the drive train needs to be cleaned, as shown in Figure 10(b). ARMAR-6 understands this message using its ASR system and recognizes that its help with the cleaning task is needed. Concretely, ARMAR-6 infers that cleaning requires a cleaning agent and the spray bottle on the nearby table, which the robot needs to localize and graspe to hand it to the technician, who has already mounted a ladder to access the drive train.

To hand over the spray bottle to the human, ARMAR-6 observes the technician, again using 3D human pose tracking. This action is shown in Figure 9 as a 3D visualization of the robot's working memory and an overlay skeleton on the input image of the RGB-D sensor.

Once the technician extends a hand toward the robot, ARMAR-6 recognizes the activity and initiates the handover. If the technician withdraws the hand, ARMAR-6 will also stop the handover and continue observing the human. However, if the technician takes the bottle from the robot's hand, it senses this with its 6D F/T sensors and opens the hand to complete the handover action. Once the cleaning task is finished, the technician will hand the spray bottle back to the robot. Again, the handover intention is recognized, and the robot's hand is closed as soon as contact is detected. While the technician stows the ladder, the robot moves back to the table and returns the spray bottle to its initial location.

Key Aspects

Figure 10 indicates only the primary sensor modality used in each action. We want to emphasize that most of the perception systems, from laser-based navigation to ASR and 3D visual perception, are constantly active, providing the robot with rich information for situational awareness and scene understanding. Throughout the entire scenario, the robot recognizes on multiple occasions that the technician needs its help (illustrated by the red circles in Figure 10). The automatic RoNoH enables the human and robot to work together naturally.

We have executed system validation studies using the previously mentioned scenario at different locations and with more than 20 different technicians—most notably in our lab at the Karlsruhe Institute of Technology, at the 2018 international CEBIT trade fair in front of large audiences, and at an actual automated fulfillment center in the United Kingdom. The usability and impact of ARMAR-6 for the maintenance tasks in this fulfillment center have been evaluated in a user study using an accompanying system usability scale questionnaire, conducted in collaboration with the SecondHands project team members of the École polytechnique fédérale de Lausanne. In this study, technicians were also asked for their subjective perception of ARMAR-6 as a coworker using the Godspeed Questionnaire Series. Details of the user study are described in [38]. The result of the evaluation helps us to continuously improve on-board algorithms for more intuitive human robot interaction.

A few of ARMAR-6's additional abilities, not showcased in the aforementioned demonstration scenario, are motion planning in unknown and dynamic environments, grasping of unknown objects, and the effortless teach-in of difficult motions using gravity-compensated, zero-torque control of the arms.

Conclusions

In this article, we provided a comprehensive overview of the current capabilities of ARMAR-6, as well as the vision we are pursuing in our research. We presented the requirements-driven design and implementation of ARMAR-6 as a platform for collaborative robotics and the design choices we made toward this goal. The joint development of hardware and software with the common goal of creating a robot that "pushes the envelope" for seamless physical HRI has led to a highly integrated, versatile, and robust humanoid robot system. The frequent demonstration of ARMAR-6 in its intended scenario, where it helps a human technician with repairing a conveyor system of an automated warehouse, not only in the laboratory but also at exhibitions and at an actual industrial warehouse, highlights its capabilities, robustness, and adaptability.

We have only just begun to explore the possibilities for research on human–robot collaboration that ARMAR-6, as an integrated robot platform, facilitates. Among other research avenues, we are currently developing methods to further improve the robot's situational awareness and human intention recognition by means of integrating multimodal sensory input, prior knowledge, and an episodic memory that allows the recalling of previous experiences. The ability to learn from human observation and experience is essential; thus, we are transferring and extending our previous work in this area to foster intuitive robot programming. Despite the robust execution of the complicated tasks described in this article, failure will occur in different situations. As a result, we are working on methods for failure prediction and recovery on the different abstraction layers of the underlying functional architecture. We are continuously upgrading the robot's hardware (e.g., improved hands and improved neck actuators) and the low-level controllers as we gain new insights into requirements through extensive use of the robot. Finally, we are working on improving the robot's dexterous manipulation skills, with a focus on grasping and manipulating unknown objects, e.g., by leveraging simulation-based learning.

Acknowledgments

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation program under grant 643950 (SecondHands). We would like to acknowledge the help and support of all other members and students of the H^{2}T lab as well as all members of the SecondHands project team, especially Aude Billard and Mahdi Khoramshahi from the École polytechnique fédérale de Lausanne; Stefan Constantin, Sebastian Stüker, and Alex Waibel from the Karlsruhe Institute of Technology; Fiora Pirri from the Sapienza University of Rome; Lourdes Agapito from University College London; and Graham Deacon from Ocado Technology. Special thanks to Michael Suppa from Roboception for providing us with one of the earliest ARMAR-6 sensors.

References

[1] A. Albu-Schäffer, S. Haddadin, C. Ott, A. Stemmer, T. Wimböck, and G. Hirzinger, "The DLR lightweight robot: Design and control concepts for robots in human environments," *Ind. Robot*, vol. 34, no. 5, pp. 376–385, 2007.

[2] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An integrated humanoid platform for sensory-motor control," in *Proc. IEEE/RAS Int.Conf. Humanoid Robots (Humanoids)*, 2006, pp. 169–175.

[3] C. Ott et al., "A humanoid two-arm system for dexterous manipulation," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2006, pp. 276–283.

[4] N. Kashiri, J. Malzahn, and N. G. Tsagarakis, "On the sensor design of torque controlled actuators: A comparison study of strain gauge and encoder-based principles," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 1186–1194, 2017.

[5] O. Stasse et al., "TALOS: A new humanoid research platform targeted for industrial applications," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2017, pp. 689–695.

[6] A. Bolotnikova et al., "A circuit-breaker use-case operated by a humanoid in aircraft manufacturing," in *Proc. IEEE Conf. Automation Science and Engineering (CASE)*, 2017, pp. 15–22.

[7] K. Kaneko et al., "Humanoid robot HRP-5P: An electrically actuated humanoid robot with high power and wide range joints," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1431–1438, 2019.

[8] M. Fuchs et al., "Rollin' Justin—design considerations and realization of a mobile platform for a humanoid upper body," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2009, pp. 4131–4137.

[9] J. Englsberger et al., "Overview of the torque-controlled humanoid robot toro," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots (Humanoids)*, 2014, pp. 916–923.

[10] N. G. Tsagarakis et al., "WALK-MAN: A high-performance humanoid platform for realistic environments," *J. Field Robot.*, vol. 34, no. 7, pp. 1225–1259, 2017.

[11] T. Yoshiike et al., "Development of experimental legged robot for inspection and disaster response in plants," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2017, pp. 4869–4876.

[12] N. Kashiri et al., "CENTAURO: A hybrid locomotion and high power resilient manipulation platform," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1595–1602, 2019.

[13] O. Khatib et al., "Ocean one: A robotic avatar for oceanic discovery," *IEEE Robot. Autom. Mag.*, vol. 23, no. 4, pp. 20–29, 2016.

[14] N. A. Radford et al., "Valkyrie: NASA's first bipedal humanoid robot," *J. Field Robot.*, vol. 32, no. 3, pp. 397–419, 2015.

[15] T. Asfour et al., "The Karlsruhe Armar humanoid robot family," in *Humanoid Robotics: A Reference*, A. Goswami and P. Vadakkepat, Eds. Dordrecht, The Netherlands: Springer, 2017, pp. 1–32.

[16] N. Vahrenkamp, M. Wächter, M. Kröhnert, K. Welke, and T. Asfour, "The robot software framework ArmarX," *it-Inform. Technol.*, vol. 57, no. 2, pp. 99–111, 2015.

[17] T. Asfour et al., "ARMAR-6: A collaborative humanoid robot for industrial environments," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2018, pp. 447–454.

[18] S. Rader, L. Kaul, P. Weiner, and T. Asfour, "Highly integrated sensor-actuator-controller units for modular robot design," in *IEEE Int. Conf. Advanced Intelligent Mechatronics (AIM)*, 2017, pp. 1160–1166.

[19] T. Asfour, "Sensorimotor coordination of movement for action execution of a humanoid robot," Ph.D. dissertation, Dept. of Informat., Univ. of Karlsruhe, Germany, 2003.

[20] S. Rader, L. Kaul, H. Fischbach, N. Vahrenkamp, and T. Asfour, "Design of a high-performance humanoid dual arm system with inner shoulder joints," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2016, pp. 523–529.

[21] N. Fukaya, S. Toyama, T. Asfour, and R. Dillmann, "Design of the TUAT/Karlsruhe Humanoid Hand," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2000, pp. 1754–1759.

[22] T. Yoshikawa, "Manipulability of robotic mechanisms," *Int. J. Robot. Res.*, vol. 4, no. 2, pp. 3–9, 1985.

[23] T. Asfour et al., "ARMAR-4: A 63 DOF torque controlled humanoid robot," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2013, pp. 390–396.

[24] N. Vahrenkamp, M. Wächter, M. Kröhnert, K. Welke, and T. Asfour, "The robot software framework ArmarX," *Inform. Technol.*, vol. 57, no. 2, pp. 99–111, 2015.

[25] D. Shingarey, L. Kaul, and T. Asfour, "Torque-based velocity control for safe human-humanoid interaction," in *Proc. 28th Int. Conf. Robotics Alpe-Adria-Danube Region (RAAD)*, 2019, pp. 61–68.

[26] Y. Zhou, M. Do, and T. Asfour, "Coordinate change dynamic movement primitives—a leader-follower approach," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2016, pp. 5481–5488.

[27] J. Gao, Y. Zhou, and T. Asfour, "Projected force-admittance control for compliant bimanual tasks," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2018, pp. 607–613.

[28] N. Vahrenkamp, E. Koch, M. Wächter, and T. Asfour, "Planning high-quality grasps using mean curvature object skeletons," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 911–918, 2018.

[29] P. Schmidt, N. Vahrenkamp, M. Wächter, and T. Asfour, "Grasping of unknown objects using deep convolutional neural networks based on depth images," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2018, pp. 6831–6838.

[30] N. Vahrenkamp, L. Westkamp, N. Yamanobe, E. E. Aksoy, and T. Asfour, "Part-based grasp planning for familiar objects," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2016, pp. 919–925.

[31] S. Ottenhaus, D. Renninghoff, R. Grimm, F. Ferreira, and T. Asfour, "Visuo-haptic grasping of unknown objects based on gaussian process implicit surfaces and deep learning," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2019.

[32] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. 2018. [Online]. Available: arXiv:1812.08008

[33] D. Tome, C. Russell, and L. Agapito, "Lifting from the deep: Convolutional 3D pose estimation from a single image," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2017, pp. 2500–2509.

[34] J. R. Medina, F. Duvallet, M. Karnam, and A. Billard, "A human-inspired controller for fluid human-robot handovers," in *Proc. IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, 2016, pp. 324–331.

[35] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Third Int. Conf. Learning Representations (ICLR)*, 2015. [Online]. Available: https://dblp.org/rec/bibtex/journals/corr/BahdanauCB14

[36] S. Constantin, J. Niehues, and A. Waibel, "Multi-task learning to improve natural language understanding," in *Proc. Tenth Int. Workshop Spoken Dialogue Systems (IWSDS)*, 2019.

[37] T. S. Nguyen, M. Sperber, S. Stüker, and A. Waibel, "Building real-time speech recognition without CMVN," in *International Conference on Speech and Computer*. Cham, Switzerland: Springer, 2018, pp. 451–460. doi: 10.1007/978-3-319-99579-3_47.

[38] B. Busch et al., "Evaluation of an industrial robotic assistant in an ecological environment," in *Proc. IEEE Int. Conf. Robot and Human Interactive Communication*, 2019.

[39] PAL Robotics. Accessed on: Oct. 12, 2019. [Online]. Available: http://reemc.pal-robotics.com/en/reemc/

[40] SecondHands. Accessed on: Oct. 12, 2019. [Online]. Available: https://secondhands.eu

[41] ArmarX, "ArmarX documentation." Accessed on: Oct. 12, 2019. [Online]. Available: https://armarx.humanoids.kit.edu

[42] YouTube, "ARMAR-6@CEBIT 2018—One shot video." Accessed on: Oct. 12, 2019. [Online]. Available: https://youtu.be/6AFkrGkKI7g

Tamim Asfour, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: asfour@kit.edu.

Mirko Wächter, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: mirko.waechter@kit.edu.

Lukas Kaul, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: lukas.kaul@kit.edu.

Samuel Rader, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: rader@kit.edu.

Pascal Weiner, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: pascal.weiner@kit.edu

Simon Ottenhaus, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: simon.ottenhaus@kit.edu.

Raphael Grimm, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: raphael.grimm@kit.edu.

You Zhou, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: you.zhou@kit.edu.

Markus Grotz, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: markus.grotz@kit.edu.

Fabian Paus, High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany. Email: paus@kit.edu.