Safe Robot Reflexes: A Taxonomy-based Decision and Modulation Framework

Jonathan Vorndamme, Alessandro Melone, Robin Kirschner, *Student Member, IEEE,* Luis Figueredo, *Member, IEEE,* and Sami Haddadin, *Fellow, IEEE*

***Abstract*—Recent advances in control and planning allow for seamless physical human-robot interaction (pHRI). At the same time, novel challenges appear in orchestrating intelligent decision-making and ensuring safe control of robots. Particularly in scenarios involving unforeseen or unintended collisions, robots face the imperative of reacting judiciously to avert potential risks to humans, other robots, obstacles, or themselves. At the same time, they need to maintain focus on their primary task or be able to safely resume it. Collision detection and identification algorithms are now well-established in industry, yet complex collision reflexes have not transitioned into industrial applications beyond basic stopping reactions. Despite the introduction of numerous advanced high-performance reflex controllers over the past decades, their real-world adoption has remained a challenge. This work establishes a systematic framework to address that gap. For this, the *reflex control problem* is defined, *reflex behaviors* are systematically classified and categorized, and relevant *safety data* is acquired following *existing international standards*. We argue that this foundational step is crucial for improving the safety and capabilities of robots in both complex industrial and domestic environments. We validate our approach within the system class of articulated manipulators through a state-of-the-art cooperative pick-and-place task, providing a blueprint for future implementations for other robot classes.**

***Index Terms*—Robot Collision Handling, Robot Reflexes, Safety, Reflex Context Classification**

Figure 1. The abrupt movement of the human to grasp the tool results in an unavoidable collision in a close human-robot collaboration scenario. Consequently, a collision reflex has to be activated to ensure the safety of both the human and the robot.

I. Introduction

As expressed by Asimov's first law - "A robot may not injure a human being or, through inaction, allow a human being to come to harm" - the safe integration of robots as co-workers in future factory settings or everyday service roles demands special care. Indeed, the proximity of robots to humans in unstructured environments forecasts (unforeseen) inevitable collisions during regular operation, see Fig. 1.

The course of such a collision is outlined in Fig. 2. In the *pre-collision* phase (a), the likelihood and severity of injury can be mitigated by factors such as robot design [1], [2], robot posture [3], and trajectory planning which collectively work to limit the robot effective mass and velocities at impact time to safe values [4], [5]. During *Impact: Phase I* (b) of the collision, the speed of the event leads to an open-loop-like behavior, with the robot- and human-reflected dynamics governing the force and energy exchange. Hence, the contact force cannot be actively reduced by the robot control system. Effective measures include lightweight robot design, soft visco-elastic covers for the links, or compliance in the drive system [6]. Instead, *Impact: Phase II* (c) is characterized by slower quasi-static contact dynamics [7], allowing low-latency control strategies to decrease the probability and severity of injury. In the *post-collision* phase (d), the robot needs to terminate the reaction and return to the nominal task.

This work was supported by the European Union's Horizon 2020 research and innovation programme as part of the project I.AM. under Grant agreement ID: 871899 and the Lighthouse Initiative Geriatronics by StMWi Bayern (Project X, grant no. IUK-1807-0007// IUK582/001) and LongLeif GaPa gGmbH (Project Y). The authors also acknowledge the support from the Lighthouse Initiative KI.FABRIK funded by StMWi, (Phase 1: Infrastructure as well as the research and development program under, grant no. DIK0249) and the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden.

The authors are with Technische Universität München (TUM), TUM School of Computation, Information and Technology (CIT), Department of Computer Engineering, Munich Institute of Robotics and Machine Intelligence (MIRMI), 80992 Munich, Germany and the Centre for Tactile Internet with Human-in-the-Loop (CeTI). Sami Haddadin, Jonathan Vorndamme, Alessandro Melone and Robin Kirschner are also with TUM Chair of Robotics and Systems Intelligence and Luis Figueredo is with the School of Computer Science, University of Nottingham, UK. Luis Figueredo is also an Associate Member at the MIRMI, Technical University of Munich (TUM). Sami Haddadin is also with Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE.

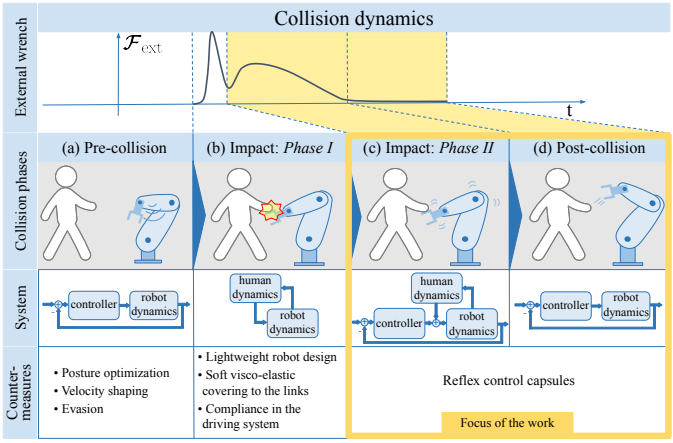


Figure 2. Illustration of the progressive collision phases in collaborative human-robot environments highlighting the collision dynamics based on the external wrenches and corresponding collision phases and system model which in turn reflect the potential counter-measures.

The problem of safety during *Impact: Phase I* has been extensively researched and is already addressed in real-world applications in the form of ISO/TS 15066:2016 [8]. Therein, force and pressure thresholds to ensure human safety for impact and clamping scenarios are defined based on prior research on the human onset of pain [4], [7], [9]. As during impact the reactive forces or pressures cannot be controlled [10], the desired thresholds can only be adhered to by velocity-scaling prior to the impact. However, the *Impact: Phase II* and *Post-collision* phases can be influenced through control mechanisms, which is therefore the primary focus of this work.

Furthermore, most injuries happening in the real world – and especially fatalities [11], [12] – are not caused by the initial impact force during *Impact: Phase I* but by the work that force exerts on the human body during the collision in *Impact: Phase II*. In an attempt to prevent these high crushing forces resulting in severe, irreversible injury, rapid safety stops have become the paramount and the only standardized collision reaction [13]. Using this collision reaction, also the post-collision robot motion never had to be considered. Nevertheless, as the focus in human-robot interaction shifts towards the prevention of even small, reversible injuries and light-weight robots being applied, the induced energy during contact needs to be reduced by any means [4], [8]. This is only possible by considering different approaches for collision reaction, potentially reducing contact times and their effect on potential secondary contacts.

In this context, robot reflexes (see Sec. II for a formal definition) play a crucial role in ensuring human safety [14]. Under laboratory conditions advanced algorithms for fast and task-consistent collision reactions already improve performance and safety in human-robot interaction [6], [15]–[20]. However, despite these technological breakthroughs in modern robot control and machine learning[1], these solutions are not yet deployed in real-world applications. Collaborative human-robot settings are classified as safety-critical [8], which implies that any robot with the potential to harm a human operator or user must be certified as safe for all operational modes in real-world settings before even being allowed to operate, in particular necessitating a thorough task- and setup-specific risk assessment around these novel methods.

Under current laws and standards, "unforeseen events" are in fact excluded and considered invalid operations. Allowing such events would be equivalent to negligence, exposing robot manufacturers and distributors to severe legal consequences. Notably, there is currently no comprehensive methodology for designing and selecting reflexes that complies with the machinery directive [21], ISO 10218 [22] and ISO/TS 15066 [8]. As a result, advanced reflex control solutions have remained confined to research laboratories, with their translation to real-world applications and commercial products still unrealized.

In this work, our primary goal is to bridge this gap by introducing a collision reflex framework that facilitates the design and analysis of collision reactions in compliance with international safety standards for robots operating in real-world environments. Our approach systematically integrates robot reflexes into the application design, risk assessment, risk reduction, and deployment phases of human-robot collaboration settings. We assume that the necessary state information and control actions are available according to the required safety rating derived from a task-specific risk assessment [23].

This work significantly enhances the autonomy of robot systems outside traditional industrial settings. In contrast to fixed collision reactions, as set by the current standards, that may ensure safety in one situation but could even pose new hazards in the next situation, our approach is strictly case-aware. This case-awareness enables robot autonomy outside of traditional settings, which enhances robotics applications as such. For example, while the stop reflex may meet safety criteria in some situations, it can introduce unnecessary risks, such as clamping, in others. Our approach evaluates individual scenarios based on existing risks and mitigation strategies, resulting in reflexes that are certifiable as *safety functions* according to [13] by design. We validate our approach through real industrial use cases (see sec. IV as well as appendix B and E in the supplementary material), demonstrating its potential impact.

Thus, in this work – in a bid to standardize safety assessments across tasks integrating diverse collision scenarios and suitable reactions – we introduce the *Robot Safety Assessment Pipeline* (RSAP), a methodology for systematic task-specific risk assessment and reduction for robot collisions according to ISO 12100 [23] (appendix A in the supplemental material describes, how the RSAP integrates into risk assessment and reduction according to ISO 12100). Our framework serves as a template for evaluating robot reflex reactions during unforeseen collisions while considering various hazards during *Impact: Phase II* and *Post-collision* phases. We demonstrate and validate the RSAP with ten carefully designed reflexes for comprehensive collision safety across diverse real-world collision scenarios.

Following this methodology yields certifiable reflex control solutions given a collaborative human-robot application. The

[1]Until now, the application of machine learning to robot reflexes is not yet possible in accordance with actual safety standards due to the lack of safety guarantees and the potentially life-threatening danger resulting from unaware reflex selection not adhering to a proper safety-guided process.

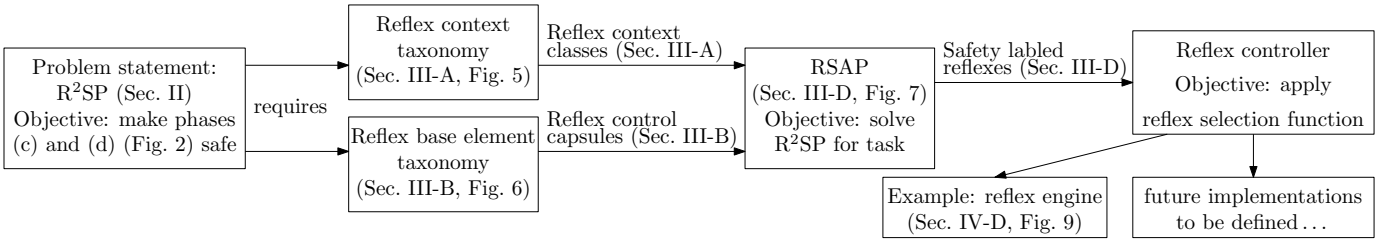


Figure 3. Workflow of the paper.

following theoretical contributions are part of the RSAP.

1) **Robot Reflex Schedule Problem (R^2SP):** defines the robot reflex and the problem of using reflex behaviors to mitigate collision hazards.
2) **Reflex Context Taxonomy:** evaluates and classifies the existing collision situations. The taxonomy is clustered in reflex context classes based on the modalities relevant for the reflex selection.
3) **Reflex Base Element Taxonomy:** defines a large space of possible reflexes to choose from, which can be easily extended by other existing works and the research community at large.
4) **Safety Vector:** defines quantities that directly measure and correlate to a given safety limit (typically according to international standards) given a concrete reflex context.

In addition to the theoretical contributions above, our work validates the RSAP in different industrial application scenarios. The innovation-oriented contributions are highlighted below:

1) A complete example of applying the RSAP to the relevant use-case of *collaborative pick-and-place* is provided.
2) The RSAP is applied to another industrial use-case (dual-arm depalletizing) featured in the European project I.AM until the point of designing reflex reference experiments (details can be found in the supplementary material, appendix B).
3) Exemplary *Safety Vectors* were deduced for the risks of *skin shearing*, *stabbing*, *cutting*, *clamping*, *unmet object affordances*, and *secondary collisions* for industrial use-cases.
4) To quantify the aforementioned *Safety Vectors*, we developed a series of standardized *Reflex Reference Experiments*. The outcome of each experiment provides a quantitative measure of safety for the specific reflex under examination.

Figure 3 illustrates the workflow of the paper. Section II defines the R^2SP which, in a nutshell, concerns ensuring safety during collision phases (c) and (d), see Fig. 2. The solution requires *reflex context classes* given by the *reflex context taxonomy* (Sec. III-A) as well as *reflex control capsules* built by the *reflex base elements* defined in the *reflex base elements taxonomy* (Sec. III-B) and the description of the task at hand. The RSAP procedure, solving the R^2SP, is explained in Sec. III-D. The application of the RSAP (Sec. IV), which includes the implementation of our exemplary reflex engine and the evaluation of real-world *reflex reference experiments* designed during the RSAP-process (Sec. V), determines whether a specific reflex can safely address a given *reflex context class* that may arise during the task under examination. The findings are analyzed in the context of current research (Sec.VI) in Sec.VII. Sec. VIII presents the concluding remarks of this paper.

II. Research Problem

This section describes the high-level system model considered in this paper, the research questions we aim to answer and a formal definition of the research problem at hand.

A. High-Level System Model

The system class (see Fig. 4) considered in this work is a joint torque-controlled robot with the desired torque input

$$\boldsymbol{\tau}_\mathrm{d} = \boldsymbol{f}_\mathrm{c}(\boldsymbol{u}, \boldsymbol{y}, t) \; . \tag{1}$$

This system operates within the closed-loop dynamics

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{x}, t) \; , \tag{2}$$

where $\boldsymbol{u}$ denotes the control input, $\boldsymbol{x}$ signifies the system state and $\boldsymbol{y}$ is the system measurement. The quantity $\boldsymbol{G}$, as shown in Fig. 4, represents the global goals (e.g. draw a line with the pen currently grasped), $\boldsymbol{g}$ denotes the policy goals (e.g. go to position A and apply a force of $1\,\mathrm{N}$ to draw the line), $\boldsymbol{S}$ indicates the global state, and t symbolizes time. The control input $\boldsymbol{u}$ encompasses the tactile policy[2], incorporating the desired motion $\boldsymbol{x}_\mathrm{d}$ of the end-effector or desired joint motion $\boldsymbol{q}_\mathrm{d}$, alongside the desired wrench $\boldsymbol{\mathcal{F}}_\mathrm{d}$. It also integrates the controller type and parameters.

A task planner delineates global goals for the skill planner, which further dissects them into force/motion goals. These are then processed by the policy planner, culminating in a tactile policy $\boldsymbol{u}$ for the robot. Additionally, a global state observer furnishes all relevant information necessary for task planning and the reflex engine (as discussed in Sec. III-A).

The reflex engine offers reactions to unforeseen collisions by transitioning the control strategy $\boldsymbol{u}_\mathrm{c}$ to reflex mode $\boldsymbol{u}_\mathrm{r}$ upon collision detection. The following distinction between reflex and control/planning is used throughout this work.

Definition 1 (Reflex): A pre-defined reaction to an unforeseen collision that instantaneously modulates or switches the control input $\boldsymbol{u}$ for a limited duration after a collision to prevent human injury and damage to the robot or the environment.

[2]A tactile policy refers to the reference trajectory generation for coordinated control of position/velocity and force in robotic manipulation tasks. This coordination is exemplified in operations such as drilling, where the robot must simultaneously maintain a specific position (motion), apply a directional force, and execute a translational movement along the force vector.

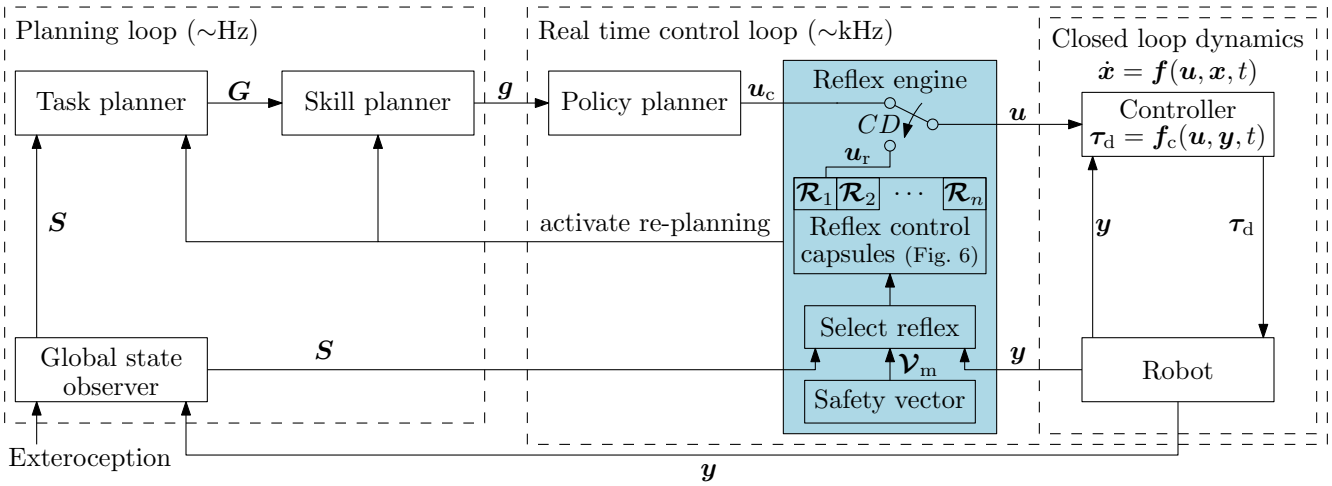


Figure 4. System class considered in this work: During normal operation, the robot controller receives inputs from the skill planner that responds to the goals from the global task planner. In case of unforeseen collisions (binary collision detection signal $CD = 1$), the reflex engine running in the real-time control loop at kHz level takes control over the robot to assure a locally safe behavior. When the critical situation is resolved, control over the system is returned to the task, skill and policy planner.

Definition 2 (Control and Planning): For a desired task, a planner provides a tactile policy with an (initially) collision-free desired trajectory, typically for the robot end-effector or joints. This is tracked and stabilized by a given real-time control strategy.

Let us exemplify this concept.

a) Example: Considering a binary collision detection signal CD and numbered *reflex context classes* $\boldsymbol{\mathcal{C}}_\mathrm{rc}$ (see sec. III-A), the resulting desired torque may be determined as follows:

$$\boldsymbol{\tau}_\mathrm{d} = \begin{cases} \boldsymbol{J}^\mathrm{T}(\boldsymbol{\Lambda}\ddot{\boldsymbol{x}}_\mathrm{d} + \boldsymbol{D}_\mathrm{d}\dot{\boldsymbol{e}} + \boldsymbol{K}_\mathrm{d}\boldsymbol{e} + \boldsymbol{\mathcal{F}}_\mathrm{d}) + \boldsymbol{C}\dot{\boldsymbol{q}} + \boldsymbol{\tau}_\mathrm{g}, & \text{if } CD{=}0 \\ \boldsymbol{\tau}_\mathrm{g}, & \text{if } CD{=}1, \boldsymbol{\mathcal{C}}_\mathrm{rc}{=}1 \\ \boldsymbol{J}^\mathrm{T}(-\boldsymbol{D}_\mathrm{d}\dot{\boldsymbol{x}} + \boldsymbol{K}_\mathrm{d}(\boldsymbol{x}(t_\mathrm{cd}) - \boldsymbol{x})) + \boldsymbol{C}\dot{\boldsymbol{q}} + \boldsymbol{\tau}_\mathrm{g}, & \text{if } CD{=}1, \boldsymbol{\mathcal{C}}_\mathrm{rc}{=}2 \\ \quad\vdots\,, \end{cases} \tag{3}$$

This exemplary control structure handles different collision scenarios. In the absence of collisions ($CD = 0$), it executes a Cartesian impedance controller with a desired feed-forward wrench $\boldsymbol{\mathcal{F}}_\mathrm{d}$. Subsequently, for distinct *reflex context classes*, specific reactions are applied, such as a *zero-g* reaction for $\boldsymbol{\mathcal{C}}_\mathrm{rc} = 1$, Cartesian based *stop* reaction for $\boldsymbol{\mathcal{C}}_\mathrm{rc} = 2$, and so forth. Above, the symbols $\boldsymbol{J}$, $\boldsymbol{\Lambda}$, $\boldsymbol{x}_\mathrm{d}$, $\boldsymbol{x}$, $\boldsymbol{D}_\mathrm{d}$, $\boldsymbol{K}_\mathrm{d}$, $\boldsymbol{C}$, $\boldsymbol{q}$, $\boldsymbol{\tau}_\mathrm{g}$, and t_cd represent the analytic Jacobian, Cartesian inertia matrix, desired Cartesian pose, current Cartesian pose, desired damping matrix, desired stiffness matrix, centrifugal and Coriolis terms, joint positions of the robot, gravitational torque, and the time of contact detection, respectively. Moreover, $\boldsymbol{e} := \boldsymbol{x}_\mathrm{d} - \boldsymbol{x}$ denotes the difference between the desired and actual Cartesian pose.

Of course, any other collision control (i.e., controllers designed to tolerate certain classes of collisions and implicitly handle them by the imposed behavior as opposed to using contact reflexes) can be applied analogously.

B. Research Questions

The primary objective of this work is to develop a reflex engine capable of identifying the right control input $\boldsymbol{u}$ that guarantees safety for humans, the robot, and the environment in all conceivable collision scenarios relevant to the task at hand. This leads to the following research inquiries:

- Q 1) What constitutes the pertinent context surrounding a collision, and which methodologies can be employed to classify various collision situations effectively with respect to choosing suitable reflexes?
- Q 2) What array of reflexes exists to ensure human safety in the workspace, protect the robot and preserve the environment during impact and post-collision phases while enabling successful task continuation following the collision resolution?
- Q 3) How can systematic approaches be devised to discern safe reactions for potential *reflex contexts* within a given task scenario?

Throughout the paper, we answer Q 1) in Sec. III-A, Q 2) in Sec. III-B and Q 3) in Sec. III-D. The problem is formally defined in the next subsection.

C. The Robot Reflex Schedule Problem (R^2SP)

It is imperative to delineate safety precisely to formally establish the *Robot Reflex Schedule Problem* (R^2SP). The ISO/TS 15066 outlines safety criteria for a collision involving a robot and a human, stipulating that the collision is safe if both the maximum quasi-static and transient contact forces remain below thresholds, contingent upon the body part impacted. While this criterion primarily focuses on the *impact phase* (phase b), see Fig. 2) and might not encapsulate the entire safety spectrum, including post-collision hazards, we adopt the premise that certain (measurable) variables can verify the safety of a reflex. This leads to the following assumption.

Assumption 1: There exists a collection of safety-defining variables constituting the safety vector $\boldsymbol{\mathcal{V}}$ and a corresponding safety set $\boldsymbol{T}$ such that the condition "$\boldsymbol{\mathcal{V}} \in \boldsymbol{T}$ at any moment during the execution of a reflex" implies that the reflex aligns with a predefined set of safety requirements, thereby ensuring safety.

The specified safety requirements may encompass objectives such as preventing human injury, safeguarding the robot from damage, or protecting objects within the environment, among others. The *Robot Reflex Schedule Problem* (R^2SP) is then formally defined as follows.

Definition 3 (Robot Reflex Schedule Problem): The objective of the *Robot Reflex Schedule Problem* entails determining a reflex schedule function $\boldsymbol{h}$ such that, given the reflex context $(\boldsymbol{y}, \boldsymbol{S})$, the function fulfills

$$\boldsymbol{u} = \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{S}) \Rightarrow \boldsymbol{\mathcal{V}} \in \boldsymbol{T} \tag{4}$$

This definition posits that the derived *reflex schedule function* $\boldsymbol{h}$ given context $(\boldsymbol{y}, \boldsymbol{S})$, ensures that the resulting action $\boldsymbol{u}$ adheres to the pre-defined safety criteria. In this study, we propose an empirically-driven approach to define the *reflex schedule function*, relying on data-supported methodologies, therefore consistent with given standards and an important step towards allowing certified robot reflexes complying to international safety regulations. The reflex theory required to solve the problem is presented in the following section.

III. Reflex Theory

A. Reflex Context Taxonomy

In prior research [2], we outlined five principal categories of human-robot collisions, namely, (i) *free*, (ii) *constrained*, (iii) *partially constrained*, (iv) *clamping in the robot structure*, and (v) *secondary impacts*.

To elucidate, *free collisions* denote instances where the collided limb is not subject to any constraints in its movement. Conversely, *constrained collisions* involve scenarios where the affected limb encounters restriction, potentially leading to clamping. The category of *partially constrained collisions* refers to incidents where the collided body part is free in its movement while other parts of the body are constrained. For example, a situation where the robot collides with the chest of a human while a box behind the person may prompt a subsequent *secondary collision* with the ground.

In terms of hazard assessment and mitigation strategies, we have established that the hazard characteristics and means of minimizing them between *clamping in the robot structure* and *constrained collisions* are largely similar, as outlined in [2]. Lastly, *secondary collisions* may occur due to two distinct reasons: (a) when the robot executes an active reflex that leads to contacts with another part of the human body, or (b) due to the human collision reaction, resulting in further contacts with the environment. This work explicitly considers collision classes (i) through (v-a), excluding (v-b) since this class is beyond the direct influence of the robot behavior and is thus not in the scope of this study.

Our earlier works [6], [24] revealed that the collision impact itself is largely uncontrollable, primarily dictated by the inherent physics of the impact (see Fig. 2), unless the robot is overly padded. Consequently, the influence of various reactions, including active *retractions*, on the peak impact force and pressure is generally insignificant, provided these reactions do not exacerbate the contact by pushing. The key factors governing the maximum impact force and pressure are the robot collision speed and its reflected mass [25], [26]. Hence, employing an evidence-grounded safety map, as detailed in [27], becomes crucial in averting human injury by imposing safe operating velocities.

Moreover, optimizing the robot posture aids in diminishing reflective mass, consequently lowering the impact force or even enabling higher operational speeds deemed safe [28]. While the choice of reaction might not significantly impact the high-speed collision phase directly, it remains crucial to mitigate post-collision hazards, particularly concerning residual clamping forces, aligning with prevailing safety standards [8]. In unconstrained contacts, stopping the robot at the collision position proves beneficial as it allows the robot to maintain its trajectory, enabling seamless continuation once the situation resolves. However, in constrained contacts, halting the robot movement could potentially induce clamping. Instead, employing a *zero-g* reflex facilitates the operator liberation. To address hazards during the post-collision phase, we introduce the *reflex context taxonomy*, depicted in Fig. 5, outlining a hierarchical tree structure for reflex contexts based on the following underlying assumptions:

Assumption 2: A single robot operates in the considered workspace.

Assumption 3: The workspace is shared with a single human.

Assumption 4: The robot experiences at most one unexpected contact with the human at any given time.

Assumption 5: The human intention for physical interaction and the collision types are known by design (from intended use and risk analysis) or an internal robot identification routine accurately computes collision signatures and classification signals, revealing them.

Assumption 6: If the necessary semantic information to determine pertinent states, encompassing the human status, potential object-related hazards in the environment, object affordances, etc., are not known by design, an *integrated context observer* possesses the ability to collect, analyze, and provide them. This observer can be an integral component of the robot or a hybrid system combining robot data, environmental knowledge, and sensory input into a unified representation.

Assumptions 2-4 significantly reduce the complexity of the reflex context without notably limiting its range of applications. The term "multiple robots" is reserved for independently controlled units; thus, a coordinated dual-arm system would still be considered a single robot. Finally, regarding multiple humans, while they could be accounted for by taking the Cartesian product of their individual states, this would exponentially increase the complexity of the system state, and such configurations are not common in current and foreseeable practice yet. Assumption 4 seems stricter than it actually is, as in the real world, two simultaneous collisions almost never occur (in most cases, one occurs slightly before the other). If the first collision is handled in a way that ensures that secondary collisions are safe – which is in any case a requirement – this will inherently lead to the second collision being safe as well.

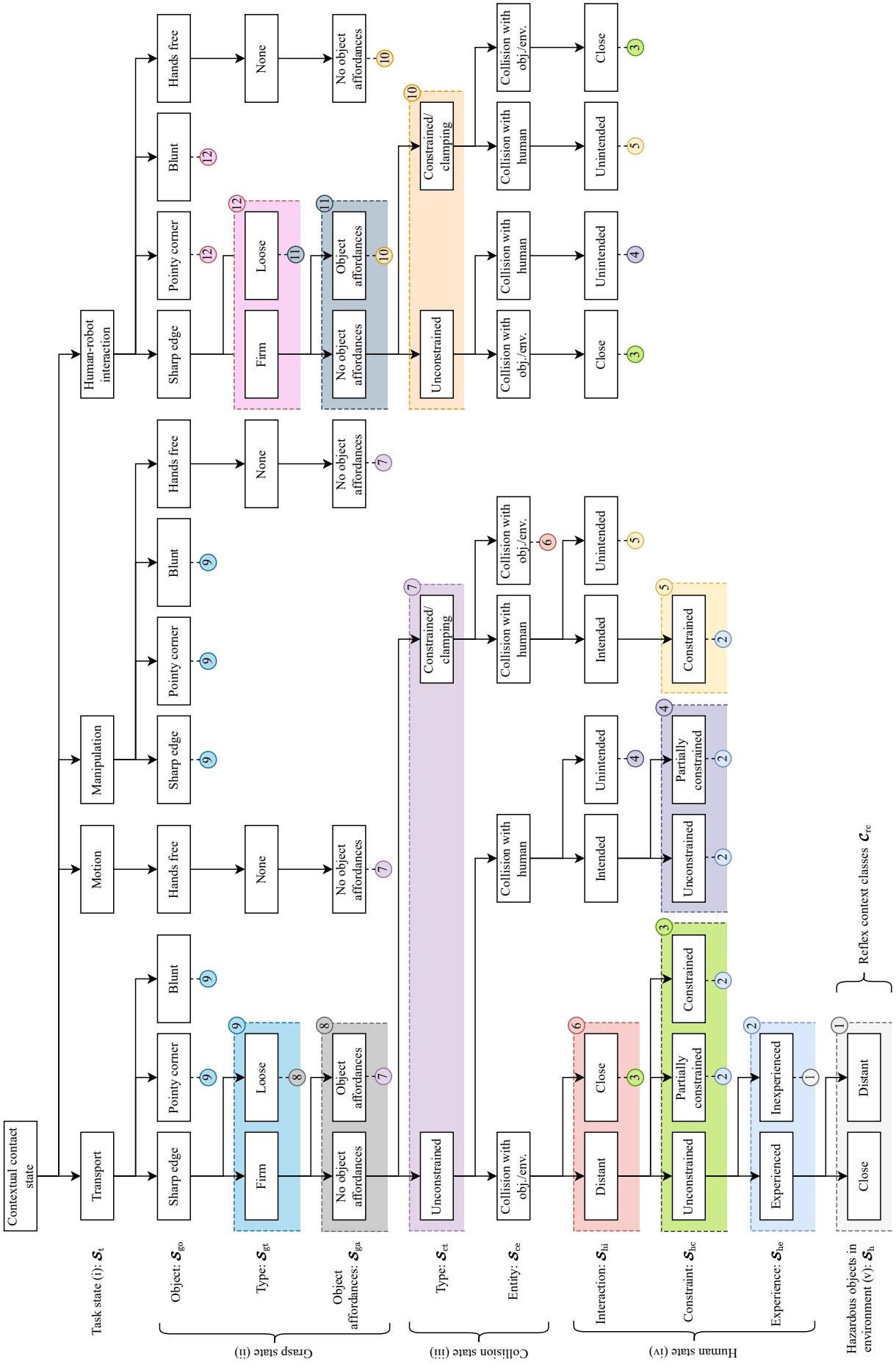


Figure 5. *Reflex context taxonomy* accommodating different task and scene conditions assembled as a taxonomy tree. Dashed boxes covering the entire branch below them represent tree branches, with branch numbers in the top-right corner. This compact notation allows branches to be efficiently referenced under other tree elements. Background colors are used to help finding the referenced subtrees. For instance, in the light blue box in the bottom left, the subtree for *inexperienced* operators is the same as for *experienced* ones (that is, number "1" containing the classification of *close* and *far away*).

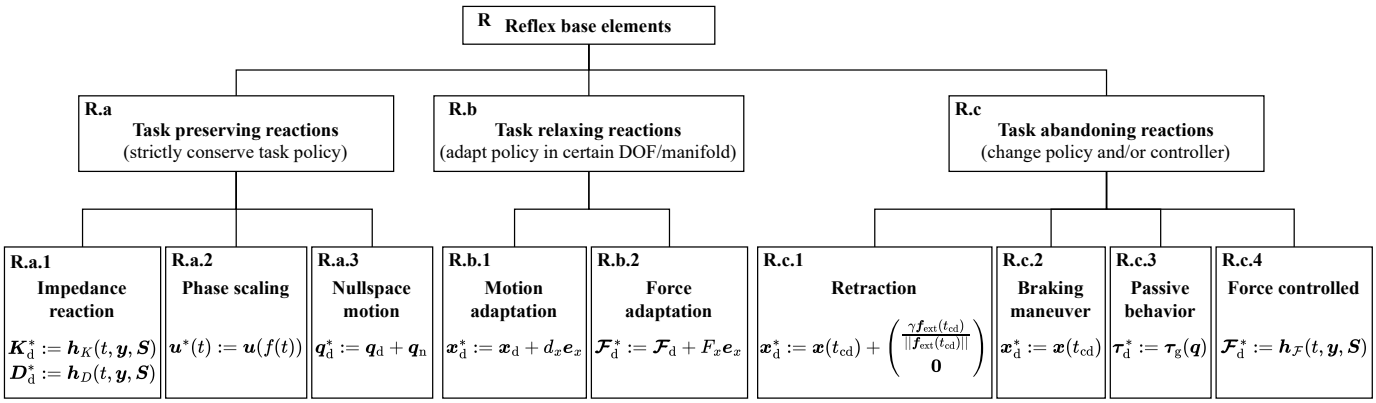


Figure 6. Reflex base element taxonomy with example algorithms. The base elements are classified into the three main categories of *task preserving*, *task relaxing* and *task abandoning* strategies. They can be combined in a timed and/or superpositional manner to obtain the set of possible *reflex control capsules*. For instance, a combination of *stop* reflex for the orientation and *zero-g* reflex for the translation could lead to an object-aware reflex that allows an open container to be moved around freely while being kept upright to avoid unwanted disposal of the contents. The asterisk denotes entities altered by the reflex. For simplification, the nominal control is assumed to be chosen, for instance, by equation (3). The symbols $\boldsymbol{q}_{\mathrm{n}}$, d_x, F_x and $\boldsymbol{e}_x$ denote a suitable nullspace displacement, the adaptation distance, the adaptation force and a 6D unit vector that is zero in the first three or the last three columns.

Despite their apparent stringency, assumptions 5 and 6 are essential for real-world application, as current safety standards prohibit "unforeseen situations" in operational contexts. Note that this does not mean that the entire state needs to be measured, as parts of it can be known by design (e.g., physical barriers could prevent constrained collisions). Still, there will be situations where the perception required for the RSAP is beyond the state of the art indicating a way for current and future safety perception research and feeding developments with concrete requirements. However, safety perception is an entire research field on its own and is clearly not the focus of this work.

The context of a reflex (see Fig. 5) is defined by (i) *task state*, (ii) *grasp state*, (iii) *collision state*, (iv) *human state* and (v) the *hazardous objects in environment* which are all part of the global state $\boldsymbol{S}$. To further delineate the *reflex context*, several key aspects are categorized and structured in the *reflex context taxonomy*:

Task state (i): This classification (symbol $\boldsymbol{\mathcal{S}}_{\mathrm{t}}$) encompasses four distinct classes. *Transport:* Involving the movement of an object from location A to location B. *Motion:* Focused on position adjustments or movements without holding an object. *Manipulation:* Involves altering an object state or the environment, possibly utilizing a tool. *Physical Human-Robot Interaction:* Close interaction between the human and robot characterized by physical contact. Task types significantly influence selecting the right reaction strategies; for instance, adjusting trajectories orthogonal to the direction of motion might be viable in transportation tasks but could risk tool damage in manipulation tasks such as drilling.

Grasp State (ii): This state is divided into the three subcategories *grasped object* $\boldsymbol{\mathcal{S}}_{\mathrm{go}}$, *grasp type* $\boldsymbol{\mathcal{S}}_{\mathrm{gt}}$ and *grasped object affordances* $\boldsymbol{\mathcal{S}}_{\mathrm{ga}}$. *Grasped objet* classifies objects by their risk properties: *Pointy Corners*, *Sharp Edges* and *Blunt Objects* [29]. The nature of the held object can pose risks, particularly to humans if it contains sharp edges or pointy corners. The *grasp type* is categorized as *Firm Grasp*, *Loose Grasp* and *None*. A loose grasp can possibly reduce the severity of harm in collisions with pointy corners if it sufficiently reduces the contact force with the pointy corner. On the other hand, a loose grasp could also lead to slippage and an object falling in a dangerous manner subsequently. The *affordances of the grasped Object* considers object affordances, such as handling a cup of hot coffee to prevent spillage, thereby avoiding harm to humans or damage to the robot.

Collision State (iii): Subcategorized into *collision type* $\boldsymbol{\mathcal{S}}_{\mathrm{ct}}$ and *collision entity* $\boldsymbol{\mathcal{S}}_{\mathrm{ce}}$. The *collision type* is classified as either *Unconstrained* or *Constrained/Clamping*. Reactions to stop the robot may lead to non-transient contacts in constrained/clamping collision cases while proving to be advantageous in unconstrained collisions. The *collision entity* identifies the entity the robot collides with, be it a *human* or an *object/environment*. The former imposes more conservative safety constraints.

Human State (iv): Categorized into three sub-states: *human interaction* $\boldsymbol{\mathcal{S}}_{\mathrm{hi}}$, *human constraint* $\boldsymbol{\mathcal{S}}_{\mathrm{hc}}$ and *human experience* $\boldsymbol{\mathcal{S}}_{\mathrm{he}}$. The *human interaction* captures whether a collision between the human and the robot was intended or unintended by the human. If the collision entity is not a human, the state captures whether the human is *close* enough to potentially encounter secondary collisions from a reflex motion or *distant* enough to avoid them. The *human constraints* refer to the movement constraints of the human and can be either *unconstrained*, *partially constrained* (the human cannot move freely in the environment, e.g., an object behind the human could cause tripping in case of being startled), or *constrained* (e.g., the hand rests on a table). The *human experience* rates the experience level into *experienced/inexperienced*, which is an important information that can be used to estimate possible human reactions to robot motions. This binary distinction is used for simplicity for now, but may be extended in the future.

Hazardous Objects in environment (v): Indicates whether objects that could harm the human in case of a secondary collision, e.g., with sharp objects or hot liquids are *close* or *distant* enough to not pose a threat. The symbol for this state is $\boldsymbol{\mathcal{S}}_{\mathrm{h}}$.

Overall, the *reflex context taxonomy* provides discrete *reflex context classes*. In addition to the information from the present *reflex context class*, measurements such as contact wrenches, poses, and velocities of the robot and the human are used for parameterization and feedback control in the reflexes.

B. Taxonomy of Reflex Base Elements

1) Reflex Classes: Given the wide range of possible reactions to unforeseen collisions, detailing each option individually becomes impractical. To streamline, we devised the *reflex base element taxonomy*, see Fig. 6. It categorizes reflexes into three primary classes: *task-preserving*, *task-relaxing*, and *task-abandoning*. Using the planned Cartesian trajectory $\boldsymbol{x}_\mathrm{d}(t)$, the planned wrench $\boldsymbol{\mathcal{F}}_\mathrm{d}(t)$, the planned joint space trajectory $\boldsymbol{q}_\mathrm{d}(t)$ and the planned applied torque $\boldsymbol{\tau}_\mathrm{a,d}$ as well as the respective versions altered by a reflex $\boldsymbol{x}^*_\mathrm{d}(t)$, $\boldsymbol{\mathcal{F}}^*_\mathrm{d}(t)$ and $\boldsymbol{q}^*_\mathrm{d}(t)$, $\boldsymbol{\tau}^*_\mathrm{a,d}(t)$, the categories can be defined as follows.

Definition 4 (R.a. task preserving reflex): A reflex $\boldsymbol{x}^*_\mathrm{d}(t)$ and $\boldsymbol{\mathcal{F}}^*_\mathrm{d}(t)$ in Cartesian space or $\boldsymbol{q}^*_\mathrm{d}(t)$ and $\boldsymbol{\tau}^*_\mathrm{a,d}(t)$ in joint space is task preserving if a continuous monotonically increasing function $f : \mathbb{R} \to \mathbb{R}$ exists such that either

$$\boldsymbol{x}^*_\mathrm{d}(t) = \boldsymbol{x}_\mathrm{d}(f(t)) \ \wedge \ \boldsymbol{\mathcal{F}}^*_\mathrm{d}(t) = \boldsymbol{\mathcal{F}}_\mathrm{d}(f(t)), \quad \forall t \in \mathbb{R} \tag{5}$$

for the Cartesian case or

$$\boldsymbol{q}^*_\mathrm{d}(t) = \boldsymbol{q}_\mathrm{d}(f(t)) \ \wedge \ \boldsymbol{\tau}^*_\mathrm{a,d} = \boldsymbol{\tau}_\mathrm{a,d}(f(t)), \quad \forall t \in \mathbb{R} \tag{6}$$

for the joint case holds.

Definition 5 (R.b. task relaxing reflex): A Cartesian reflex $\boldsymbol{x}^*_\mathrm{d}(t)$ and $\boldsymbol{\mathcal{F}}^*_\mathrm{d}(t)$ is called task-relaxing if for $i \in \{1, \dots, 6\}$ there exists an orthonormal matrix $\boldsymbol{T} \in \mathbb{R}^{6\times6}$ and $j \in \{1, \dots, 5\}$ such that

$$\begin{aligned} &\forall i \leq j, \ \forall t \in \mathbb{R} : \ (y^*_i(t) = y_i(t) \wedge F^*_i(t) = F_i(t)) \\ &\wedge \forall i > j : \ (\exists t \in \mathbb{R} : \ y^*_i(t) \neq y_i(t) \vee \mathcal{F}^*_i(t) \neq F_i(t)) \end{aligned} \tag{7}$$

where $\boldsymbol{y}(t) := \boldsymbol{T}\boldsymbol{x}_\mathrm{d}(t)$, $\boldsymbol{y}^*(t) := \boldsymbol{T}\boldsymbol{x}^*_\mathrm{d}(t)$, $\boldsymbol{F}(t) := \boldsymbol{T}\boldsymbol{\mathcal{F}}_\mathrm{d}(t)$ and $\boldsymbol{F}^*(t) := \boldsymbol{T}\boldsymbol{\mathcal{F}}^*_\mathrm{d}(t)$. A joint space reflex $\boldsymbol{q}^*_\mathrm{d}(t)$ and $\boldsymbol{\tau}^*_\mathrm{a,d}(t)$ is called task relaxing if for $i \in \{1, \dots, n\}$ there exists an orthonormal matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ and $j \in \{1, \dots, n-1\}$ such that

$$\begin{aligned} &\forall i \leq j, \ \forall t \in \mathbb{R} : \ (p^*_i(t) = p_i(t) \wedge m^*_i(t) = m_i(t)) \\ &\wedge \forall i > j : \ (\exists t \in \mathbb{R} : \ p^*_i(t) \neq p_i(t) \vee m^*_i(t) \neq m_i(t)) \end{aligned} \tag{8}$$

where $\boldsymbol{p}(t) := \boldsymbol{T}\boldsymbol{q}_\mathrm{d}(t)$, $\boldsymbol{p}^*(t) := \boldsymbol{T}\boldsymbol{q}^*_\mathrm{d}(t)$, $\boldsymbol{m}(t) := \boldsymbol{T}\boldsymbol{\tau}_\mathrm{a,d}(t)$, $\boldsymbol{m}^*(t) := \boldsymbol{T}\boldsymbol{\tau}^*_\mathrm{a,d}(t)$ and n is the number of joints.

Definition 6 (R.c. task abandoning reflex): A reflex is categorized as task-abandoning if it neither falls into any of the above classes. *Remark:* Task abandoning reflexes include all torque level reflexes.

Appendix C in the supplementary material elaborates on the interpretation of the above definitions.

2) Reflex Subclasses: In our previous studies [6], [24] and various other existing literature, such as [15], [19], [30]–[40], the reactions categorized as task-abandoning can be further organized into three main classes: *braking*, *passive behavior*, and *retraction*. Each of these exhibits distinct characteristics and implications.

Basic braking reflexes without considering the environment often induce clamping, while active *retraction* reflexes can lead to secondary collisions if the robot lacks awareness of the retraction space. Passive behaviors, on the other hand, such as zero-g or admittance control, mitigate the risk of subsequent clamping if not entirely eliminating it. However, they do not completely eradicate the possibility of secondary collisions. This risk may be further mitigated by enhancing compliance, for instance, adopting *super-zero-g* (see Sec. IV-C) instead.

These strategies heighten the probability of task failure, as the response must consider the affordances of objects held by the robot. For instance, a container might need to be held upright to prevent an inadvertent spill. The *reflex base elements taxonomy* further delineates all categories into more detailed subclasses, see Fig. 6.

In principle, braking maneuvers can be implemented such that they do not alter the trajectory of the robot, so one could assume them to be a task preserving strategy. However, in this work, we assume them to be task abandoning as the task is only continued on user input and thus the robot is not able to finish the task on its own.

The set of all possible reactions considered in this work is derived from the base elements by combination in a timed (e.g. *zero-g* for 2 seconds, then *stop*) and/or superpositional manner through geometric primitives (e.g. gravity free translation while orientation is held constant). Such a ready-to-apply combination of reflex base elements is called a *reflex control capsule*, see Fig. 6.

C. Reflex Selection Function

1) Definition: The *reflex space* concept aims to encompass all potential assignments of *reflex control capsules* derived from *reflex base elements* to *reflex context classes* outlined in the *reflex context taxonomy*. The *reflex selection function* r is structured as follows (see Table I for notation):

$$r : \mathcal{HS} \times \mathcal{GS} \times \mathcal{WS} \times \mathcal{CS} \to \mathcal{C} \times \mathcal{P} \times \mathcal{G} \times \mathcal{T}, \tag{9}$$

This function combines semantic insights from the *reflex context taxonomy* with a quantitative assessment of the scene, such as poses and velocities of objects in the workspace, the human and the robot, as well as contact wrenches. It leads to selecting a concrete *reflex control capsule* that ensures safety within the prescribed context. The selected *reflex control capsule* is then executed using a sequence of parameterized controllers to achieve the desired trajectory and goal, see Sec. IV-E.

A *reflex selection function* r may be implemented on a subset of the $\mathcal{HS} \times \mathcal{GS} \times \mathcal{WS} \times \mathcal{CS}$ set. However, situations that fall outside its defined domain should be avoided in this case. In practical scenarios, only a limited subset of reflex context classes is typically relevant. The *reflex space* encompasses the collective set of all *reflex selection functions* capable of ensuring safety across all situations within their respective domains.

In this work, the reflex selection is made based on the empirically grounded safety performance, i.e., the reflex with

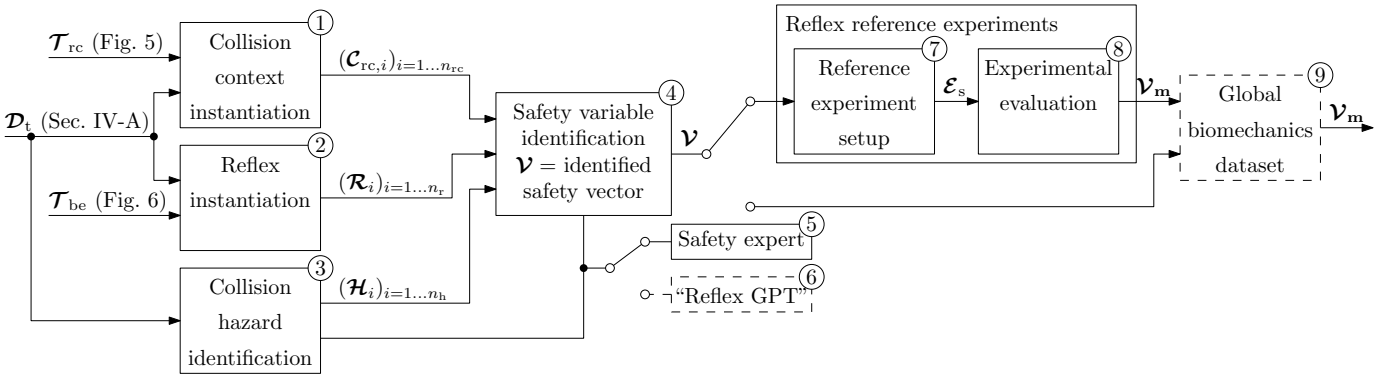


Figure 7. Robot Safety Assessment Pipeline (RSAP). Based on the task description $\mathcal{D}_{\mathrm{t}}$, the reflex context taxonomy $\mathcal{T}_{\mathrm{rc}}$ and the reflex base element taxonomy $\mathcal{T}_{\mathrm{be}}$, the possibly occurring reflex context classes $\mathcal{C}_{\mathrm{rc},i}$, candidate reflexes $\mathcal{R}_i$ and possibly occurring hazards $\mathcal{H}_i$ are determined. They are used to deduce the safety vector $\mathcal{V}$. If the global biomechanics dataset does not have a safety vector measurement for the reflex in question yet, a reference experiment setup $\mathcal{E}_{\mathrm{s}}$ is designed to obtain the safety vector measurement $\mathcal{V}_{\mathrm{m}}$ which is subsequently stored in the dataset. In the future, having a dataset of safety vector measurements accumulated over time will allow skipping the measurement and query the safety vector $\mathcal{V}_{\mathrm{m}}$ from the dataset. Furthermore, gathering and leveraging the increasingly large human injury protection data will allow an AI agent (exemplary called "reflex GPT" here) to algorithmically substitute the safety expert in the composition of $\mathcal{V}$.

Table I
EXPLANATION OF REFLEX SELECTION FUNCTION SYMBOLS.

Symbol	Entity	Description
$\mathcal{HS}$	human state	limb/body part/head poses, twists, contact, experience
$\mathcal{GS}$	grasp state	object affordances (e.g. keep upright), sharp/pointy edges, reflected mass
$\mathcal{WS}$	world state	obstacles poses/velocities/constraints
$\mathcal{CS}$	contact state	wrench-direction, wrench magnitude, location, link, constrained
$\mathcal{C}$	controllers	sequence of controllers used for reflex execution
$\mathcal{P}$	controller parameters	sequence of parameters for above controllers
$\mathcal{G}$	goals	sequence of goals for reflex, e.g., reach a certain state, priority (such as hitting uncritical part), ...
$\mathcal{T}$	trajectories	sequence of trajectories (force/torque/position/orientation) for reflex execution

the highest overall performance for the present *reflex context* is chosen. However, the most important point is not how the reflexes are chosen in the end but how to generate a set of valid reflex options guaranteeing a required margin of safety. The next section describes the pipeline that synthesizes safe reflex options.

D. Robot Safety Assessment Pipeline: Find Safe Reactions

In order to be able to select safe reflexes, a systematic method to evaluate the reflex compliance with given safety criteria is needed. The *Robot Safety Assessment Pipeline* (RSAP) structured in Fig. 7 serves as a framework for evaluating these key safety aspects of robotic tasks. Its primary goal is to identify a set of safe reflexes based on actually experimentally validated contextualized safety performance, thereby limiting the selection space to guarantee safety. In the future, and with enough such data this will allow machine learning techniques, for instance, to learn the selection of reflexes based on the reflex context while guaranteeing safety throughout the process. However, for now the reflex selection is conducted based on the overall safety performance.

The pipeline encompasses several key steps.

1) *Task Description:* Define the task and the environment, including possible hazards.
2) *Collision Hazard Identification:* Determine the collision hazards and risks associated with task execution.
3) *Collision context instantiation:* Go through the task step by step and find all possibly occurring collision contexts according to the reflex context taxonomy in Fig. 5.
4) *Reflex instantiation:* Defines the set of fully parameterized reflexes that is to be tested for applicability in the previously defined collision contexts.
5) *Safety variable identification:* For each of the previously defined risks, find the physical entities allowing the deduction of the safety towards the respective risk and the according safety thresholds. They are selected based on expert experience with experiments, see, e.g., [41], [42].
6) *Reflex reference experiments:* Design reproducible benchmark experiments to reliably measure the safety vector in well defined worst case situations, such that if the chosen reflex complies with the threshold in the experiments, it also complies in collisions occurring during the task[3].

The RSAP can be utilized in two primary ways: 1.) verify the safety of a task with predefined reflexes or 2.) identify suitable reflexes for specific reflex contexts occurring in a given task. The next section provides an example application of RSAP to a collaborative pick-and-place task (see Fig. 8).

IV. RSAP USE CASE: SYSTEM DESIGN

The subsequent sections commence with the task description, a necessary input for the RSAP, and systematically navigate

[3]Experimental validation is strictly necessary here as current safety standards do not allow for purely theoretical safety concepts.

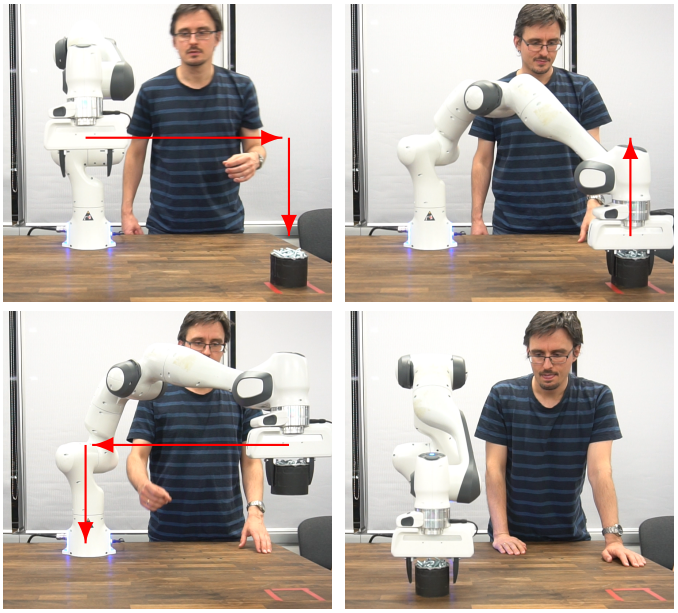


Figure 8. Example task for application of RSAP. Red arrows show the motion of the robot end-effector.

through the RSAP blocks according to Fig. 7.

A. Task Description (Fig. 7, Input)

For illustrative purposes, we elaborate on a generic safety fenceless pick-and-place task within the proposed framework, see Fig. 8. The task involves securely grasping an open container filled with screws and relocating it to a different position on a table. The robot is situated on the table alongside the container, while an expert human operator is nearby, not intending to interact with the robot. The scene contains no additional objects, constraining possible operator movements solely by the table, the robot, and the container. Both the robot and the task planner have access to the geometric scene description, ensuring that unexpected collisions are only conceivable with the human.

B. Collision Hazard Identification (Fig. 7, Block ③)

1) Impact vs. Post-Collision Hazards: As outlined in Fig. 2, the impact itself is not directly controllable, thus we assume the direct collision safety is handled, for instance, through an SMU [43]. ISO/TS 15066 sets criteria for impact safety, considering maximum contact forces $F_{1,\max}$ and transient forces $F_{1,\mathrm{qs}}$ depending on the body part involved in the collision. However, collision hazards in the form of post-impact hazards, that occur during the contact phase (*Impact: Phase II* in Fig. 2) and post-collision hazards that occur after the contact (*Post-collision* in Fig. 2), still pose significant risks and these need to be defined.

2) Considered Hazards: The hazards we focus on in this work are as follows.

a) Skin shearing: occurs when the human body part is constrained, and the robot moves orthogonally to the contact force, causing the skin to tear apart due to friction forces.

b) Stabbing: occurs if a pointy object presses onto human skin, potentially penetrating through skin or muscle tissue.

c) Cutting: occurs when sharp edges slide over the skin, potentially causing cutting wounds even at minimal force.

d) Clamping: occurs when a human body part is immobilized by the robot against an environmental constraint for an extended duration (over $0.5\,\mathrm{s}$ [8]), potentially causing severe contusions.

e) Unmet Object Affordances: are hazards caused by transported objects, e.g. falling objects or spilled hot liquids.

f) Secondary Collision: denotes the hazards caused by collisions due to active or passive evasive motions of the robot during a reflex.

C. Collision Context Instantiation (Fig. 7, Block ①)

Referring to the *reflex context taxonomy* in Fig. 5, we delineate the categorization of the *reflex context class* $\boldsymbol{\mathcal{C}}_{\mathrm{rc}}$ contingent on the corresponding contact situation.

The *task state* $\boldsymbol{\mathcal{S}}_{\mathrm{t}}$ for the fenceless pick-and-place is *motion* until the robot picks up the container, transitioning to a *transport* task afterward (see Fig. 5). During the motion task, the *grasped object* $\boldsymbol{\mathcal{S}}_{\mathrm{go}}$ and *grasp type* $\boldsymbol{\mathcal{S}}_{\mathrm{gt}}$ are naturally *hands-free* and *none*, respectively. Given that the robot carries no object with affordances, the *grasped object affordances* $\boldsymbol{\mathcal{S}}_{\mathrm{ga}}$ are *no object affordances*. Possible collision types $\boldsymbol{\mathcal{S}}_{\mathrm{ct}}$ could be *unconstrained* or *constrained/clamping*. For example, a constrained collision occurs when the human hand is located between the robot and the container at the time of collision.

Given that the robot is aware of the environment and all objects in the scene, the *collision entity* $\boldsymbol{\mathcal{S}}_{\mathrm{ce}}$ will be *collision with human*. The *human interaction state* $\boldsymbol{\mathcal{S}}_{\mathrm{hi}}$ is invariably *unintended*, as there is no intention for interaction. Furthermore, the *human experience state* $\boldsymbol{\mathcal{S}}_{\mathrm{he}}$ is *experienced*, and the human constraint state $\boldsymbol{\mathcal{S}}_{\mathrm{hc}}$ is *constrained* in *constrained/clamping* contacts, whereas it is *unconstrained* for *unconstrained* contacts, as no other objects are present that could lead to a *partially constrained* $\boldsymbol{\mathcal{S}}_{\mathrm{hc}}$. The *hazardous objects in the environment* $\boldsymbol{\mathcal{S}}_{\mathrm{h}}$ are *distant*, as no such objects are present.

During the transportation task, the *grasped object* $\boldsymbol{\mathcal{S}}_{\mathrm{go}}$ is *blunt* as the container has no sharp edges or corners. The *grasp type* $\boldsymbol{\mathcal{S}}_{\mathrm{gt}}$ is defined as *firm*, and the *grasped object affordances* $\boldsymbol{\mathcal{S}}_{\mathrm{ga}}$ are *object affordances*, as the container needs to be held upright. Except for this, the same rationale used for the motion task applies during the transportation task regarding collision state, human state, and hazardous object state.

In summary, four potential *reflex contexts* are conceivable: *motion task* and *free contact*, *motion task* and *constrained/clamping contact*, *transportation task* and *free contact*, as well as *transportation task* and *constrained contact*.

Before continuing with the next step of RSAP, the *reflex instantiation*, we describe the reflex engine that is responsible for selecting reflexes and integrating them into the task flow. The RSAP will label all tested reflexes into suitable or unsuitable for all collision situations identified in this step such that the reflex engine, when applied in the task, has a set of suitable reflexes to select from for all occurring collision situations. The selection itself can then be done heuristically, based on task or safety performance, or perspectively even via machine learning based classification without compromising safety.

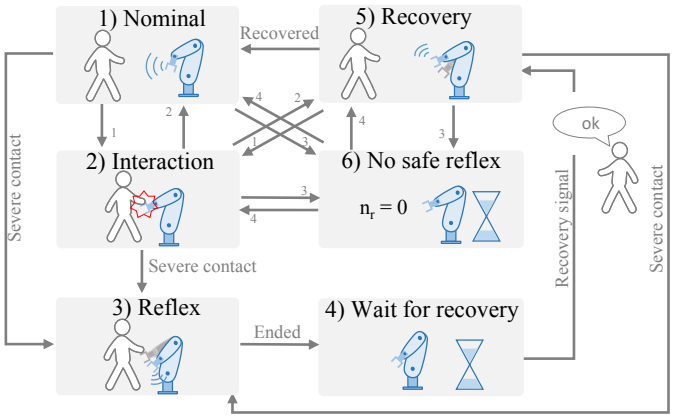


Figure 9. Reflex State machine[5]. In the event of light contact, a *task preserving* or *task relaxing* reflex is executed, e.g., slowing down the robot. On severe or repeated contacts, a *task abandoning* reflex that requires a recovery motion afterward is executed, such as switching control to zero-g mode. After the reflex ends, it waits for a signal (e.g., haptic interaction by the user). Then the robot aims to recover the task at hand. Suppose the currently selected reflex cannot handle possibly occurring collisions safely (i.e., the *reflex selection function domain* (RF-domain) is left); the reflex engine enters the out of RF-domain state, where the robot is stopped and waits until the domain of the reflex engine is re-entered again either by re-planning from the global planner or by removal of the obstacle that causes the hazard that cannot be avoided safely.

D. Reflex Engine: Design and Implementation

Figure 4 illustrates the intricate interplay among control, skill planning, and task planning within our reflex framework. While various design methodologies exist for reflex engines [44], our approach adopts a state machine model, see Fig. 9. The following subsections outline the states typically traversed during a task, presenting their sequence.

1) Nominal: The *Nominal* state signifies the passive observation phase. During this phase, the system state is actively monitored, and the reflex engine remains on standby for potential collision events. It assesses external forces using the robot state and stays informed about the global state through the global state observer. Light and severe contacts trigger transitions to the interaction and reflex states, respectively.

2) Interaction: The *Interaction* state reflects task-preserving and task-relaxing reactions to light contacts, such as stiffness adaptation. If the contact ends, the state machine reverts to the previous state. In case of a severe contact, the *Reflex* state is entered.

3) Reflex: Task-abandoning reactions like zero-g control are implemented in the *Reflex* state to ensure safety during collision processes.

4) Wait for recovery: After completing the reflex execution, the robot awaits a recovery signal, which could be initiated by various means such as a timeout, user input, program-issued signal, or haptic interaction.

5) Recovery: In the *Recovery* state, the robot aims to recover the task by retracing steps to the collision position, resuming the task from that point. This conservative approach is based on the assumption that the task and motion planner originally devised a collision-free task considering all static objects in the scene. On successful reflex execution, there is a high probability that the path from the current location to the collision location is free, as the robot already successfully traversed this space. However, the pathway from the current location to the original goal position might remain obstructed. Therefore, for prioritizing reliability, we sacrifice potential task progress during reflex execution. Once the collision position is reached, the robot transitions back to normal operation and resumes the task execution.

6) No safe reflex: The *No safe reflex* state is entered when the current global state resides outside the domain of the current *reflex selection function*. This scenario implies that an imminent collision cannot be resolved safely, prompting the robot to halt before the collision unfolds. Resumption of the task execution occurs when the global state re-enters the domain of the *reflex selection function*. This re-entry might be due to a new *reflex selection function* offering a safe response for the current state, alterations in the environment (e.g., movement of a human previously obstructing the robot path), or the introduction of a new task plan that circumvents the potentially hazardous situation altogether.

[5]It is essential to note that while the diagram portrays a state machine, it simplifies the true complexity. The states of interaction and out-of-domain in a real state machine necessitate memory of their previous states to accurately return when exited. However, for clarity and compactness, this complexity is streamlined in the diagram. This can be resolved by establishing multiple instances of the mentioned states, each connected to only one of the normal operation and recovery states.

Table II
CANDIDATE REFLEXES FOR RSAP USE CASE

Name	Control law	Parameters
Stop	$\boldsymbol{q}_{\mathrm{d}}^{*}(t) = \boldsymbol{q}(t_{\mathrm{cd}})$	
Admit-tance	$\dot{\boldsymbol{q}}_{\mathrm{d}}^{*}(t) = \alpha \boldsymbol{\tau}_{\mathrm{ext}}(t)$	$\alpha = 0.05$
Zero-g	$\boldsymbol{\tau}_{\mathrm{d}}(t) = \boldsymbol{\tau}_{\mathrm{g}}(t)$	
Super-zero-g	$\boldsymbol{\tau}_{\mathrm{d}}(t) = \boldsymbol{\tau}_{\mathrm{g}}(t) - \beta \boldsymbol{\tau}_{\mathrm{ext}}(t)$	$\beta = 0.5$
Hybrid	$x_{\mathrm{d},i}^{*}(t) = x_i(t_{\mathrm{cd}}) \quad \forall i \in \{4,5,6\}$ (orientation DOF)	
Cart-retract	$\boldsymbol{x}_{\mathrm{d}}^{*}(t) = \boldsymbol{x}(t_{\mathrm{cd}}) + \begin{pmatrix} \frac{\gamma \boldsymbol{f}_{\mathrm{ext}}(t_{\mathrm{cd}})}{\|\boldsymbol{f}_{\mathrm{ext}}(t_{\mathrm{cd}})\|} \\ \mathbf{0} \end{pmatrix}$	$\gamma = 0.09$
Joint-retract	$\boldsymbol{q}_{\mathrm{d}}^{*}(t) = \boldsymbol{q}(t_{\mathrm{cd}}) + \frac{\delta \boldsymbol{\tau}_{\mathrm{ext}}(t_{\mathrm{cd}})}{\|\boldsymbol{J}_{\mathrm{Cart}} \boldsymbol{\tau}_{\mathrm{ext}}(t_{\mathrm{cd}})\|}$	$\delta = 0.05$
Stop-retract	$\boldsymbol{q}_{\mathrm{d}}^{*}(t) = \begin{cases} \boldsymbol{q}(t_{\mathrm{cd}}), & t < t_{\mathrm{cd}} + \epsilon \\ \boldsymbol{q}(t_{\mathrm{cd}}) + \frac{\delta \boldsymbol{\tau}_{\mathrm{ext,max}}}{\|\boldsymbol{J}_{\mathrm{Cart}} \boldsymbol{\tau}_{\mathrm{ext,max}}\|}, & t \geq t_{\mathrm{cd}} + \epsilon \end{cases}$	$\delta = 0.05$ $\epsilon = 0.1$
Admit-tance-stop	$\dot{\boldsymbol{q}}_{\mathrm{d}}^{*}(t) = \begin{cases} \alpha \boldsymbol{\tau}_{\mathrm{ext}}(t), & t < t_{\zeta} \\ 0, & t \geq t_{\zeta} \end{cases}$	$\alpha = 0.05$ $\zeta = 0.02$
Super-stop	$\boldsymbol{\tau}_{\mathrm{d}}(t) = \boldsymbol{\tau}_{\mathrm{g}}(t) - \beta \boldsymbol{\tau}_{\mathrm{ext}}(t), \quad t < t_{\zeta}$ $\boldsymbol{q}_{\mathrm{d}}^{*}(t) = \boldsymbol{q}(t_{\mathrm{dist}}), \quad t \geq t_{\zeta}$	$\beta = 0.5$ $\zeta = 0.02$

E. Reflex Instantiation (Fig. 7, Block ②)

Based on existing literature and our previous work, ten candidate reflexes are selected, see Tab. II. The symbols t_{cd}, $\boldsymbol{\tau}_{\mathrm{d}}$, $\boldsymbol{\tau}_{\mathrm{g}}$, $\boldsymbol{f}_{\mathrm{ext}}$, $\boldsymbol{\tau}_{\mathrm{ext,max}}$, t_{ζ} denote the time of contact detection, the desired and gravity compensation torques, the external force on the end-effector, the maximum external torque during the

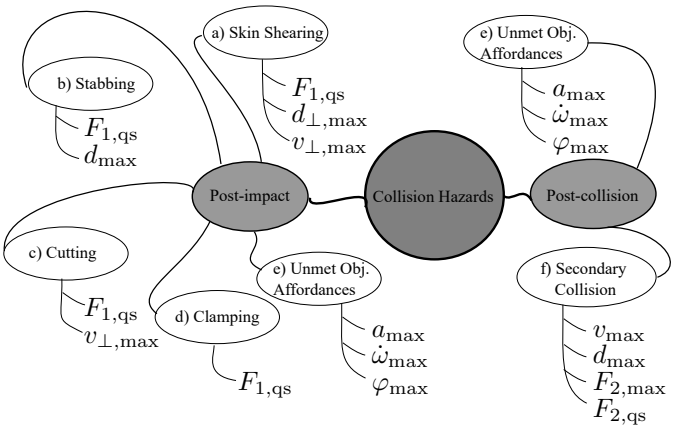


Figure 10. Taxonomy of collision hazards and assigned safety variables (by an expert, see Fig. 7). A distinction is made between post-impact hazards caused by the collision phase following the impact and post-collision hazards that may be caused be the robots reaction to the collision after the contact as been de-established.

ϵ-second stop period of the *stop-retract* reflex, and the time at which the end-effector has moved $\zeta\,\mathrm{m}$ away from the location of contact detection, respectively.

The reflexes also include *stop*, *admittance*, *zero-g*, and *super-zero-g*, from our prior work [24]. In the *stop* reflex, the controller goal position is set to the current position at the moment of collision detection. The *admittance*, *zero-g*, and *super-zero-g* reflexes switch the controller to admittance control, zero-g control, and super-zero-g control (zero-g control with inertia shaping [45]), respectively. The *hybrid* reflex is a combination of *zero-g* for translation and *stop* for orientation. In other words, the robot orientation is fixed while it can move freely in space in translation.

We also considered active retractions with joint (*joint-retract*) and Cartesian (*cart-retract*) impedance control. For Cartesian retraction, the goal position is set to $\gamma = 9\,\mathrm{cm}$ away from the current position along the direction of the external force. Regarding joint retraction, the same behavior is approximated by mapping Cartesian distance through the geometric Jacobian $\boldsymbol{J}_{\mathrm{Cart}}$, with $\delta = 5\,\mathrm{cm}$ bounds, and scaling the external torque accordingly. The difference in distance bounds is due to the lower stiffness of the Cartesian controller and the aim of designing the reflexes to accelerate away from the contact at a similar rate.

Additionally, we designed three reflex maneuvers based on the results of experiments in our previous work [24]. The *stop-retract* reflex stops the robot for $\epsilon = 100\,\mathrm{ms}$ after the collision and then retracts along the direction of the maximum force measured during the stop phase. The *admittance-stop* and *super-stop* reflexes switch to admittance and super-zero-g control, respectively, and halt the robot as soon as it moves farther than $\zeta = 2\,\mathrm{cm}$ away from the collision location. This distance was chosen such that the human has enough space to move away from the collision location comfortably but secondary collisions are unlikely. The parameters α and β where chosen to require minimal forces to move the robot around while still ensuring stability of the controller.

F. Safety Variable Identification (Fig. 7, Block ④)

The safety variables considered in this work expand on prior research [9]. Figure 10 summarizes the results detailed in the following.

a) Skin shearing: The safety vector elements are the transient, i.e. quasistatic, contact force $F_{1,\mathrm{qs}}$, the robot velocity orthogonal to the contact force $v_{\perp\max}$ and the maximum distance traveled orthogonal to the contact force during contact $d_{\perp\max}$. While $v_{\perp\max}$ and $d_{\perp\max}$ determine how fast and far the relative motion between the skin and the robot is, $F_{1,\mathrm{qs}}$ indicates the forces applied during the motion.

b) Stabbing: The safety vector elements are $F_{1,\mathrm{qs}}$ and the maximum distance $d_{\max}$ traveled into the contact. The stabbing depth is given by $d_{\max}$ and the penetration force by $F_{1,\mathrm{qs}}$.

c) Cutting: Safety vector elements are $F_{1,\mathrm{qs}}$ and the maximum velocity $v_{\perp\max}$ orthogonal to the contact force during the contact.

d) Clamping: The safety vector element is $F_{1,\mathrm{qs}}$. It is assumed that if $F_{1,\mathrm{qs}}$ is below a threshold, severe clamping is prevented.

e) Unmet Object Affordances: The safety vector elements are the maximum translational $a_{\max}$ and rotational $\dot{\omega}_{\max}$ acceleration and the maximum angular deviation from the nominal orientation $\varphi_{\max}$. High acceleration may lead to spilling [46] and a deviation in orientation could cause carried objects to fall.

f) Secondary Collision: Safety vector elements are the maximum post-collision velocity $v_{\max}$, the maximum distance traveled from the contact location during evasive motion $d_{\max}$ as well as the maximum $F_{2,\max}$ and transient $F_{2,\mathrm{qs}}$ secondary contact forces. The severity of a potential secondary collision can be estimated by $v_{\max}$, $F_{2,\max}$ and $F_{2,\mathrm{qs}}$. The forces should always be lower than the primary impact equivalents to assure safety. The likelihood of reaching another object or human body part increases with $d_{\max}$.

g) Thresholds: For the four reflex situations relevant to the considered pick-and-place task, thresholds were determined according to Fig. 11. The rationale behind these thresholds is as follows.

1) $v_{\perp\max}$ is limited according to $v_{\max}$ as no pointy or sharp objects are involved,
2) $a_{\max}$ and $\dot{\omega}_{\max}$ are only limited by the robot physical limits as no spillable liquids are considered.
3) $F_{1,\max}$, $F_{2,\max}$, $F_{1,\mathrm{qs}}$, and $F_{2,\mathrm{qs}}$ are limited according to standard thresholds for constrained or unconstrained collision with the human hand [8].
4) $d_{\max} = 0.1\,\mathrm{m}$ and $v_{\max} = 0.25\,\mathrm{m/s}$ are chosen, based on the nominal velocity.
5) We chose $\varphi_{\max} = \pi/18$ to prevent dropping the load of the container.

G. Reflex Reference Experiments (Fig. 7, Block ⑦ and ⑧)

As the final step in the RSAP, three reference experiments are proposed to objectify and measure the safety vector for the instantiated reflexes from Sec. IV-E. Details on the experiments, results, and deduced reflex performance are discussed in Section V.

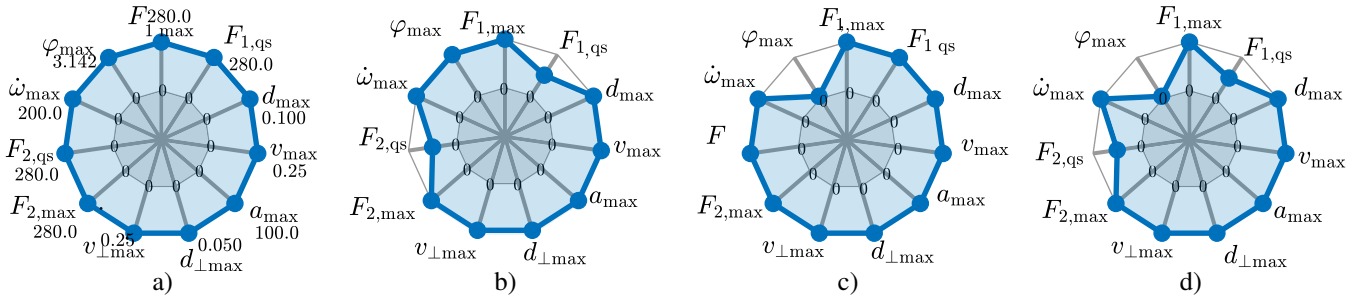


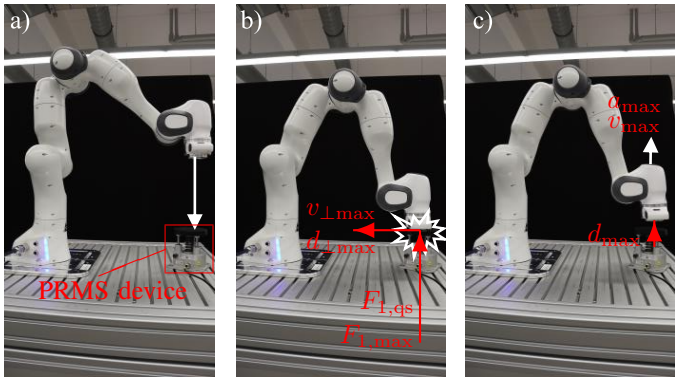
Figure 11. Thresholds for the four different *reflex context classes* occurring in the example task according to the *collision context instantiation*: free collision without object in hand (a), constrained collision without object in hand (b), free collision with object in hand (c), constrained collision with object in hand (d). The axis scaling is the same for all plots.

a) Experiment 1: Primary Collision Safety Variables: To measure $F_{1,\max}$, $F_{1,\mathrm{qs}}$, $v_{\perp\max}$, $d_{\perp\max}$, $d_{\max}$, $v_{\max}$, and $a_{\max}$, the robot collides with the PRMS device[6] [47], see Fig. 12 1). This device, a 1-DOF force sensor standardized to comply with ISO/TS 15066, simulates the collision behavior of human body parts and measures the collision force for 1 s after initial contact, applicable for our maximum and transient force criteria. Other relevant quantities are retrievable from the robot measurements.

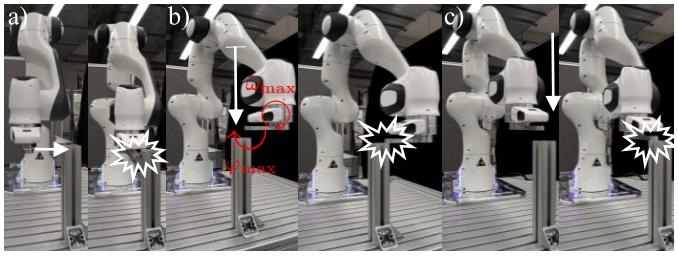
b) Experiment 2: External Moment Reactions: The reactions of reflexes are tested when external wrenches with significant moments are applied to the end-effector, see Fig. 12 2). The variables of interest are $\dot{\omega}_{\max}$ and $\varphi_{\max}$. An impactor is attached to the robot end-effector, creating a lever arm of approximately 19 cm. Three tests are conducted, each with the robot end-effector following a straight-line movement to induce torques around the three axes of the robot base coordinate system.

c) Experiment 3: Secondary Collision Severity: While primary collisions have been extensively studied due to their obvious potential to cause severe injuries, secondary collisions have received minimal attention in the literature. However, understanding various collision reflexes also involves observing the potential risks arising from subsequent movements directly after the actual collision. In our experimental framework, we exemplary revisit contact scenarios involving a human hand. The objective of the proposed experiment is to evaluate the risks associated with secondary contacts, such as the potential for high peak forces or inadvertent clamping. Hence, the PRMS-device emulating the human hand was placed in the opposite direction of the initial collision, allowing retraction actions to induce the observed secondary contacts.

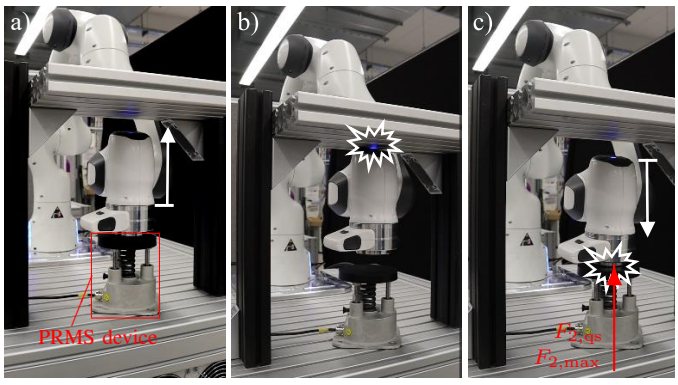
The severity of secondary collisions is assessed by measuring $F_{2,\max}$ and $F_{2,\mathrm{qs}}$. The robot collides with a static obstacle positioned above the PRMS device that obstructs the retraction space. This setup allows for evaluating potential secondary collisions caused by evasive reactions with the PRMS device, see Fig. 12 3).



1) Setup for experiment 1. The collision generates pure forces and $F_{1,\max}$, $F_{1,\mathrm{qs}}$, $v_{\perp\max}$, $d_{\perp\max}$, $d_{\max}$, $v_{\max}$ as well as $a_{\max}$ are obtained from PRMS and robot measurement data.



2) Setup for experiment 2. The collision exerts torques and forces on the end-effector and $\dot{\omega}_{\max}$ as well as $\varphi_{\max}$ are obtained from robot measurement data.



3) Setup for experiment 3. The robot collides at the top and the reaction drives it into the PRMS device at the bottom. The PRMS device measures $F_{2,\max}$ and $F_{2,\mathrm{qs}}$.

Figure 12. Experiment setups.

V. RSAP Use Case: Safety Vector Data Generation

In this section, a series of experimental trials, as described in Sec. IV-G, are presented, focusing on the evaluation of robot

[6]https://www.pilz.com/de-DE/produkte/robotik/prms/prms

reflexive capabilities in response to anticipated impact scenarios. To ensure the generalizability of these experiments, contact locations are selected based on the standardized reference cube described in EN ISO 9283:1996 [48]. Additionally, a safety test apparatus, in compliance with the forthcoming EN ISO 10218-2:2021 [22] (currently in draft stage) for precise force measurements, is employed. For consistent evaluation across different scenarios, collisions are intentionally directed towards the central reference cube location at coordinates [498, 0, 252] mm, as recommended in [49], or along the central axis of the cube closer to the table at [498, 0, 152] mm due to its anticipated significance in applications.

To establish reliable collision detection, conservative fine-tuning of thresholds has been carried out. The set thresholds, ensuring the absence of false positives, are 10 N for absolute external Cartesian force and $3\,\mathrm{Nm}$ for absolute external Cartesian torques. Primary and secondary collisions with the PRMS device, featuring a $75\,\mathrm{N/mm}$ spring and a black cover with a hardness of 70 ShA (Shore hardness scale A), are employed to assess the robot responses. All collision responses are programmed and executed on a Franka Emika Robot operating at a frequency of $1\,\mathrm{kHz}$ using the Franka Control Interface (FCI) [50].

A. Experiment 1: collision reaction to pure external forces

This scenario simulates a situation where a robot may encounter a human hand in a shared workspace, such as when a human is working close to a table. The robot, engaged in tasks like picking up or placing an object, unintentionally collides with a resting human hand.

To experimentally analyze this scenario, a PRMS device is employed, designed to mimic a human hand. The device is equipped with a $75\,\mathrm{N/mm}$ spring stiffness and a $70\,\mathrm{ShA}$ cover. Positioned centrally on a table within the robot workspace (see Fig. 12 1a), the device is set up at the starting position ($0.495\,\mathrm{m}$, $0\,\mathrm{m}$, $0.355\,\mathrm{m}$), located $240\,\mathrm{mm}$ below the robot end-effector.

The experiment commences by initiating measurements with the PRMS device, allowing for multiple autonomous repetitions. The robot descends along its end-effector z-axis, recording internal measurements. Once reaching an average speed of $0.228\,\mathrm{m/s}$, approximately the maximum speed for human-robot interaction according to [8], the robot collides with the PRMS device, executing a predefined reaction, see Fig. 12 1b) and 1c). Subsequently, the robot returns to its initial position, repeating the process for all outlined reflexes.

B. Experiment 2: collision reaction to moments (and forces)

The second experiment is designed to assess collisions when the robot holds a tool during a task, introducing a lever arm on the robot flange. In this configuration, the robot moves toward a rigid obstacle, resulting in a collision, with forces and moments monitored through the internal robot state. The obstacle is intentionally positioned at the center of the reference cube. An impactor, extending $200\,\mathrm{mm}$, is attached from the robot flange coordinate system to the side.

This collision is particularly arranged to produce significant external moments with a lever arm of $190\,\mathrm{mm}$ around the robot flange along x-, y-, and z-axes, see Fig. 12 2a), 2b), and 2c). Following the collision reaction, the robot returns to its initial position, repeating the process with subsequent collision reaction strategies.

C. Experiment 3: secondary collision severity

The experimental setup contains a rigid obstacle placed above the robot end-effector, see Fig. 12 3a). This gives the robot $\approx 3.5\,\mathrm{cm}$ of free motion space for the primary collision reflex[7]. The robot accelerates upward towards the obstacle, reaching its maximum acceleration of $1.7\,\mathrm{m/s^2}$, and eventually collides (average speed of 0.125 m/s) with the rigid profile, see Fig. 12 3b).

Depending on the collision reaction (passive, active, or no retraction), the robot is propelled downwards towards the PRMS-device, see Fig. 12 3c). This experiment is replicated for all considered collision reactions.

D. Results

The assessment pipeline involves comparing the resulting *safety vectors* without any object involved. The resulting *reflex performance spectrum* of the ten reflex schemes from Tab. II, can be found in Fig. 13. While ideal thresholds for the safety variables would be based on biomechanical data, lacking such information led to heuristic threshold selections based on experience. Deeper knowledge and a comprehensive biomechanical database that are steadily growing [4], [29] could refine these thresholds and explore additional variables. However, these aspects are clearly beyond the scope of this study.

From Fig. 13, it is evident that the *admittance* reflexes generate maximum velocities significantly surpassing the set limits. As a result, potential secondary contacts produce forces exceeding the permissible thresholds. Notably, during the experiment, the *admittance* reflexes caused the robot to surpass its dynamic limits, triggering an error state and braking before a secondary contact occurred.

For the *admittance-stop* reflex, a similar situation emerged where immediately after the secondary contact substantial quasi-static forces were generated. The *joint-retract* and *stop-retract* reflexes, utilizing similar retraction methods, exhibited slightly excessive speeds but halted before secondary collisions occurred. However, the *Cartesian-retract* reflex led to a secondary contact, producing forces below limits.

Excluding the first four reflexes mentioned above, all other reflexes appeared suitable for the considered collision situations. Among these, *stop* and *zero-g* reflexes emerged as the most effective for this scenario. The *stop* reflex naturally reduces the risk of secondary collisions or excessive motions during contact, albeit it is susceptible to clamping. On the other hand, the *zero-g* reflex demonstrated the strongest damping among the tested passive evasion reflexes, reducing the likelihood

[7]The distance was carefully selected to ensure robot acceleration up to a velocity of $0.125\,\mathrm{m/s}$ based on simulations. This is a trade-off between a collision velocity close to the maximum velocity for human-robot interaction according to [8] and a small gap between the obstacle and the PRMS device to enable secondary collisions.

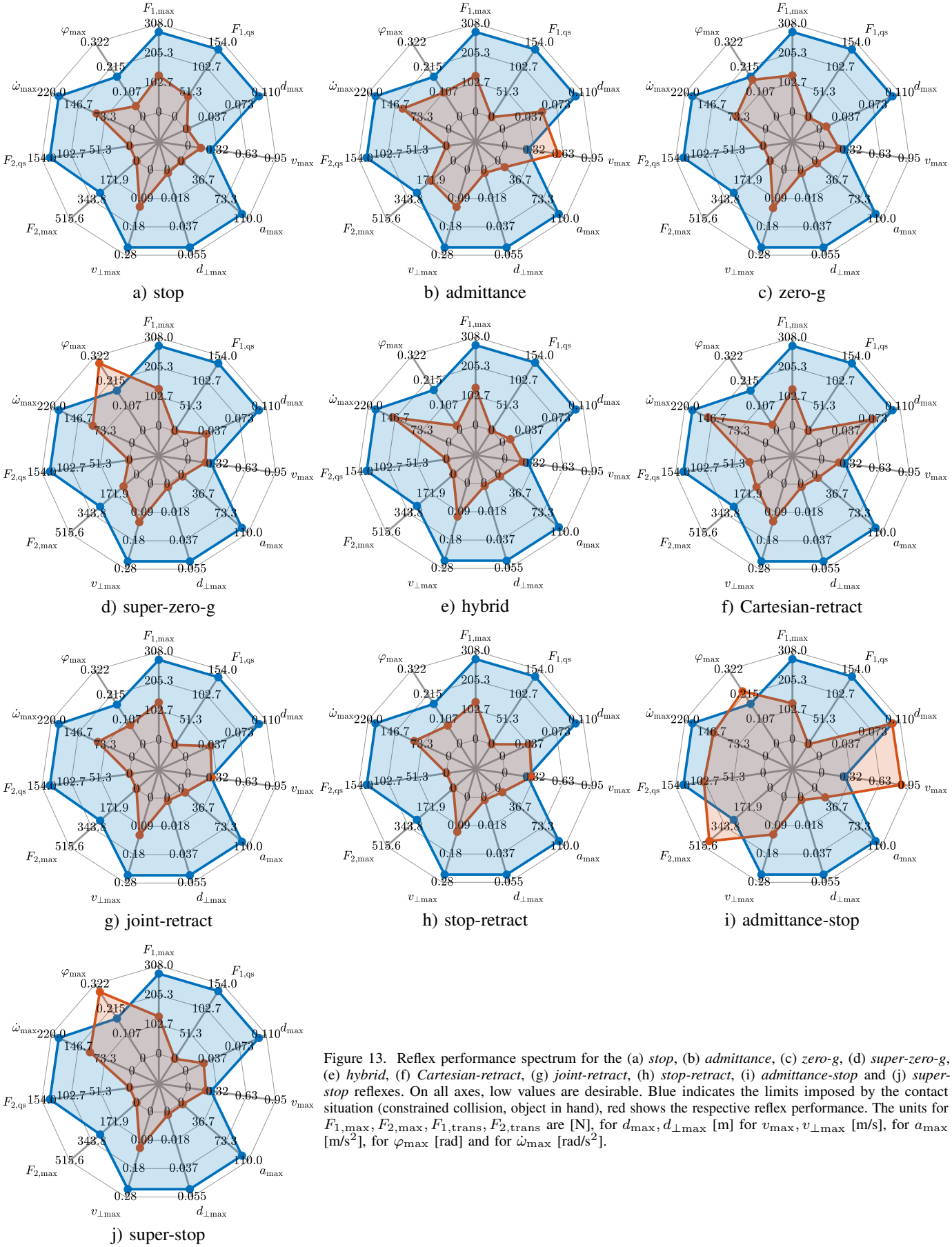


Figure 13. Reflex performance spectrum for the (a) *stop*, (b) *admittance*, (c) *zero-g*, (d) *super-zero-g*, (e) *hybrid*, (f) *Cartesian-retract*, (g) *joint-retract*, (h) *stop-retract*, (i) *admittance-stop* and (j) *super-stop* reflexes. On all axes, low values are desirable. Blue indicates the limits imposed by the contact situation (constrained collision, object in hand), red shows the respective reflex performance. The units for $F_{1,\max}, F_{2,\max}, F_{1,\mathrm{trans}}, F_{2,\mathrm{trans}}$ are [N], for $d_{\max}, d_{\perp\max}$ [m] for $v_{\max}, v_{\perp\max}$ [m/s], for $a_{\max}$ [m/s^2], for $\varphi_{\max}$ [rad] and for $\dot{\omega}_{\max}$ [rad/s^2].

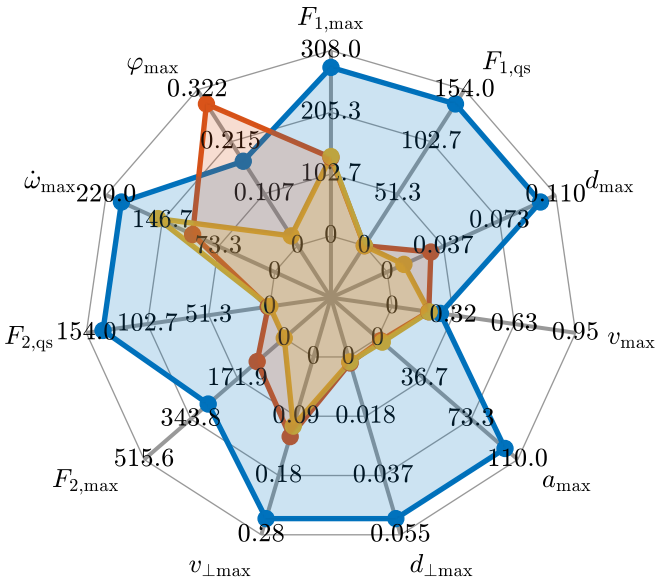


Figure 14. Comparison between *super-zero-g* (red) and *hybrid* (yellow) reflex for constrained collisions with an object in hand against the thresholds (blue).

of secondary collisions or orientations conflicting with object affordances.

Finally, Fig. 14 depicts a comparison between the *super-zero-g* and *hybrid* reflexes for constrained collisions while holding the cup. The *super-zero-g* reflex allows excessive orientation deviations, failing to meet object affordances. Conversely, as expected, the *hybrid* reflex maintains a nearly constant orientation. It might seem counter-intuitive that the *hybrid* reflex, aimed at stabilizing orientation, produces a higher maximum angular acceleration. However, this acceleration is a counter-action to the moment induced by the collision, enabling the reflex to maintain the orientation effectively.

VI. State of the Art

A. Collision Classification Schemes

The first human-robot collision classification work in [9] introduced the Safety Tree. It outlines potential injury scenarios, major factors contributing to worst-case scenarios, and their ranges for various contact situations. Building upon this, the Safe Motion Unit (SMU) [4], [43] was introduced to ensure human biomechanical safety during collisions by constraining robot velocities to impact-safe levels. Diverse schemes further classify contacts into intended interactions and unintended collisions [30]–[32], [51], [52], while others use thresholds on maximum contact force to classify collision severity [19], [24], [33], [34].

These works and several others culminated into the ISO/TS 15066:2016(E), which aims to ensure human safety in pHRI mainly by differentiating between quasi-static and transient contacts [9]. However, this standard primarily considers rigid contact surfaces larger than $1\,\mathrm{cm}^2$ for injury limits, and assumes a safety-rated monitored *stop* [13], [53] as reaction [8].

B. Robot Collision Reactions

The literature on robot collision reaction strategies offers a vast array of approaches. The most prevalent strategy remains stopping, a practice endorsed by collaborative robot manufacturers and ISO/TS 15066 guidelines [6], [15], [20], [24], [35], [54]–[56]. Additionally, admittance control offers a viable alternative which has been well-explored in literature, applied either exclusively for intended interactions, constrained within the task-specific nullspace – up to a certain threshold – or ensuring compliance to the robot's torque limits [6], [19], [24], [32], [34]–[37], [54], [55].

Variable stiffness actuators offer a unique response by reducing the stiffness while moving proportional to the estimated external torque [15]. Other studies explore active *retractions* [16], human-inspired reflexes mimicking pain responses [18], contact avoidance strategies using force vectors [17], [31], [32], or torque-/acceleration-based reactions [37].

A sensitive skin system is proposed in [35], designed to detect and isolate collisions. This system is adept at executing a variety of reactions in response to collisions, including *stopping*, *retracing* along the desired path, *retracting* to a predefined posture, applying pressure to stabilize contact, zero-g control, and slowing down in anticipation of the impact.

Another study [38] suggested employing impedance or force control as a reaction strategy to unforeseen contacts in human-robot interaction.

C. Integration and Choice of Reaction

Various methodologies exist to integrate reflexes into control frameworks and the decision-making process in order to select suitable collision reactions. In the existing literature, a well-established approach involves hierarchical integration, which first attempts to resolve the contact in the task nullspace. However, it may necessitate abandoning the task based on a predetermined contact force threshold [19], [33], [34]. Another prevalent approach involves choosing a reaction based on the classification of the contact into either intended interactions or unintended collisions [30]–[32], [51], [52].

In [31] the authors concentrate on integrating task execution and reaction mechanisms through behavior blending. Alternative approaches, such as [30], [32], utilize state machines to integrate reactions into the task execution process. Similarly, a decision tree approach is utilized in [35] to select an appropriate reaction.

D. Collision Reaction Performance Evaluation

Despite the vast body of literature on collision reaction strategies, studies that systematically evaluate the reflex performance are sparse, particularly when seeking statistical evidence.

Multiple studies analyse the maximum external force or torque for various different reflexes [6], [16], [20], [24], [39], [54], [55], [57] or joint stiffnesses and soft covers [40]. Interestingly, while in some works, different reflexes lead to significantly different peak external forces [39], in others that is not the case [16], [24], [54], [55] and some show both behaviors in some of the evaluated scenarios [6], [20], [57]. While the peak contact force is an important feature in evaluating the risk of collisions, as shown in the example in Sec. IV, it does not sufficiently describe all hazards, leaving a relevant open research question in terms of collision safety.

VII. Discussion

A. Synopsis: Safety Vector Measurement

Reflexes exhibiting a passive response to contacts, such as *zero-g* and *admittance*, demonstrate their suitability for constrained contacts due to their diminished risk of clamping. However, it is noteworthy that incorporating higher damping is advisable to mitigate more elaborate motions, thereby reducing the risk of secondary collisions. In contrast, braking maneuvers, while lacking the risk of unwanted motion, may result in clamping. *Active retractions*, falling somewhat between these extremes, exhibit limited impact and transient forces during secondary collisions. Yet, they also possess the potential to cause clamping at the secondary collision location. These findings confirm the classification of reactions into *braking*, *passive evasion*, and *active retraction*, as different implementations of these approaches showcase similar behavioral patterns. Table VI in appendix D in the supplementary material summarizes the reflex properties.

B. Reflex taxonomy

In this study, we introduce two taxonomies: the *reflex context taxonomy* and the *reflex base elements taxonomy*, both instrumental in defining the *reflex selection function*. The *reflex context taxonomy*, as depicted in Fig. 5, serves to provide comprehensive semantic information critical for selecting the most suitable reaction strategy. Unlike previous methodologies such as the "Safety Tree" [9], we emphasize on categorizing contact situations for optimal reaction strategies rather than solely focusing on potential injury threats.

Notably, our taxonomy offers a broader perspective in terms of the number of contact situations, diverging from and significantly extending established norms such as ISO/TS 15066 [58]. This expanded view is based on our assumption that the latest and future generation of collaborative [59], soft [60], and tactile [50] robots can function safely across a wider spectrum of contexts involving human proximity. While already significantly extending current norms and standards, it is straight forward to further generalize the approach to other domains, robot classes and multi-contact situations.

It is essential to recognize that if the robot velocity exceeds safe limits during impact, the resulting high contact forces in the initial impact phase (as depicted in Fig. 2) signify that no robot reflex could prevent severe human injury. Hence, if the robot velocity is deemed unsafe, our reflex engine enters the *No safe reflex* state, halting the task execution. The Safe Motion Unit (SMU) [4], [43], which underlying concept has been the basis for current standards, seamlessly integrates into our framework, aiding in avoiding unsafe velocities.

The *reflex base elements taxonomy* serves as a succinct yet comprehensive summary of validated robot reflexes. By combining reflexes from the leaf nodes, tailored *reflex control capsules* specific to addressing safety concerns in a particular scenario can be formulated. This adaptability allows for the creation of robot reflexes tuned to the specific safety requirements of the problem at hand.

C. Integration and Choice of Reaction

The proposed reflex engine integrates various prior approaches. Specifically, reflex controls that involve selecting a reaction based on contact classification can be integrated by assimilating the classification outcomes into the world state. This incorporation allows decision trees within the reaction states of the reflex engine to discern and opt for the most suitable reaction based on the context.

An example of this approach albeit without classification is presented in [24]. Moreover, the reflex engine explicitly integrates into the task workflow through the explicit handling provided by the *wait for recovery* and *recovery* states. This integration ensures that the reflex engine operates seamlessly within the larger framework of task execution and planning.

D. Future Work

The RSAP approach is conceptually comprehensive. However, some aspects remain open. First, to enable practical deployment, a comprehensive database consisting of clustered benchmark experiments and specific thresholds for all identified hazards is required. While benchmark experiments for some classes of biomechanical injuries have been convicted, detailed safety thresholds and deeper insights into the safety defining variables are required. Further exploration and definition of these thresholds are vital for effective safety assessment.

Another challenge lies in the distinction between manipulation forces and collision forces, especially when relying on proprioceptive contact detection. Discriminating between these forces solely based on proprioceptive sensors becomes challenging. The reason is that accurately distinguishing manipulation forces from collision forces demands either highly accurate models of manipulation forces or a substantial difference in magnitude between the two forces. Modelling manipulation forces robustly, especially considering variations in the timing of contact, remains a challenge [61].

In the future, the realization of a global state observer is important for the fully fledged reflex engine in real world applications to distinguish between different collision classes for reflex selection.

E. Impact

The RSAP methodology is a unifying approach that systematically expands the scope of human-robot collaboration while ensuring a higher degree of human safety. By providing a structured framework for safety assessment, RSAP mitigates the limitations imposed by existing robot safety norms, allowing for a wider range of human-robot collaborative scenarios.

Moreover, the versatility of the RSAP framework allows for its extension to more complex robotic systems such as mobile platforms or humanoid robots. This could broaden the applicability of robots in everyday life by enhancing safety and enabling their use in diverse scenarios, such as collaborative manufacturing, household service robotics (e.g. clearing the dishwasher) or assistance with tasks not in direct physical interaction with the patient in eldery- and healthcare such as making the bed or serving food and drinks (see appendix E

for a more detailed description of the tasks and a preliminary RSAP application).

While a reflex selection method based on the overall safety performance has been proposed in this work, it is important to note that the reflex selection itself is not the central aspect of this work. The main advancement of the RSAP is to systematically provide a set of safe – in the industrial and not research meaning of the word – reflex options for different *reflex contexts*, thus ensuring safety regardless of the selection method in use. This even facilitates machine learning approaches for reflex selection in the future, as traditional machine learning approaches faced challenges in guaranteeing safety during the learning phase. However, with RSAP's guidance in identifying reflexes that meet minimum safety requirements to be provided in well established deterministic technology, it becomes feasible to integrate machine learning to optimize reflex and task performance while still guaranteeing safety.

VIII. Conclusion

Engaging in a broad range of tasks alongside humans, robots inevitably carry the risk of unforeseen collisions. Choosing an appropriate collision reflex tailored to the specific collision situations is crucial for safely preventing injury. This important problem is formally introduced as the *Robot Reflex Schedule Problem* (R^2SP). The different collision types and possible reflexes are classified in the *reflex context taxonomy* and *reflex base elements taxonomy* presented in this work, respectively.

To evaluate the safety of robot reflexes, we introduce the *Robot Safety Assessment Pipeline* (RSAP). This framework formalizes the safety assessment process for robot reflexes, utilizing a safety vector – comprising safety defining variables – and reference experiments designed to measure it. Consequently, each reflex is associated with a measured safety vector, facilitating reflex selection within the reflex engine. To validate our approach, an exemplary reflex engine was implemented and successfully applied to a collaborative, fenceless pick-and-place task.

While our safety assessment framework represents a significant stride, defining the safety variables and their thresholds, along with the design of reference experiments for their measurement, remains to be done systematically. Addressing these gaps requires further fundamental research, particularly in understanding the emergence of biomechanical injuries. Additionally, practically capturing the complete state necessary to evaluate the *reflex context taxonomy* for arbitrary contact situations remains a challenging task.

The proposed framework holds immense potential in enabling closer yet safe collaboration between robots and humans, prioritizing safety without compromise. Furthermore, RSAP serves as a secure foundation for machine learning applications focused on R^2SP, heralding the development of the next generation of reflex-enabled robots.

References

[1] J.-J. Park, S. Haddadin, J.-B. Song, and A. Albu-Schäffer, "Designing optimally safe robot surface properties for minimizing the stress characteristics of human-robot collisions," in *ICRA*, 2011, pp. 5413–5420.

[2] S. Haddadin, A. Albu-Schäffer, M. Frommberger, J. Rossmann, and G. Hirzinger, "The 'DLR crash report': Towards a standard crash-testing protocol for robot safety-part i: Results," in *ICRA*, 2009, pp. 272–279.

[3] O. W. Maaroof and M. İ. C. Dede, "Physical human-robot interaction: Increasing safety by robot arm's posture optimization," in *ROMANSY 21 - Robot Design, Dynamics and Control*. Cham: Springer International Publishing, 2016, pp. 329–337.

[4] S. Haddadin, S. Haddadin, A. Khoury, T. Rokahr, S. Parusel, R. Burgkart, A. Bicchi, and A. Albu-Schäffer, "On making robots understand safety: Embedding injury knowledge into control," *IJRR*, vol. 31, no. 13, pp. 1578–1602, 2012.

[5] R. Laha, W. Wu, R. Sun, N. Mansfeld, L. F. Figueredo, and S. Haddadin, "S*: On safe and time efficient robot motion planning," in *ICRA*, 2023, pp. 12 758–12 764.

[6] A. De Luca, A. Albu-Schaffer, S. Haddadin, and G. Hirzinger, "Collision detection and safe reaction with the dlr-iii lightweight manipulator arm," in *IROS*, 2006, pp. 1623–1630.

[7] S. Haddadin and E. Croft, *Springer Handbook of Robotics, Ch. Physical Human–Robot Interaction*. Cham: Springer International Publishing, 2016, pp. 1835–1874.

[8] DIN ISO/TS 15066:2016-02, Robots and robotic devices — Collaborative robots (ISO/TS 15066:2016).

[9] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, "Requirements for safe robots: Measurements, analysis and new insights," *IJRR*, vol. 28, no. 11-12, pp. 1507–1527, 2009.

[10] S. Haddadin, A. O. Albu-Schäffer, and G. Hirzinger, "Safety evaluation of physical human-robot interaction via crash-testing," in *RSS*, 2007.

[11] B. Young. (accessed 24th of September 2024) The first 'killer robot' was around back in 1979. HowStuffWorks. [Online]. Available: https://science.howstuffworks.com/first-killer-robot-was-around-back-in-1979.htm

[12] E. Atkinson. (accessed 19th of July 2024) Man crushed to death by robot in south korea. BBC News. [Online]. Available: https://www.bbc.com/news/world-asia-67354709

[13] DIN EN ISO 13849-1:2016-06, Safety of machinery - Safety-related parts of control systems - Part 1: General principles for design (ISO 13849-1:2015).

[14] J. Vincent. (accessed 4th of July 2024) Chess robot breaks seven-year-old's finger during tournament in russia. The Verge. [Online]. Available: https://www.theverge.com/2022/7/25/23276982/chess-robot-breaks-childs-finger-russia-tournament

[15] A. De Luca, F. Flacco, A. Bicchi, and R. Schiavi, "Nonlinear decoupled motion-stiffness control and collision detection/reaction for the vsa-ii variable stiffness device," in *IROS*, 2009, pp. 5487–5494.

[16] F. Cavenago, M. Massari, A. M. Giordano, and G. Garofalo, "Unexpected collision detection, estimation, and reaction for a free-flying orbital robot," *JGCD*, vol. 44, no. 5, pp. 967–982, 2021.

[17] S. Haddadin, H. Urbanek, S. Parusel, D. Burschka, J. Roßmann, A. Albu-Schäffer, and G. Hirzinger, "Real-time reactive motion generation based on variable attractor dynamics and shaped velocities," in *IROS*, 2010, pp. 3109–3116.

[18] J. Kuehn and S. Haddadin, "An artificial robot nervous system to teach robots how to feel pain and reflexively react to potentially damaging contacts," *RA-L*, vol. 2, no. 1, pp. 72–79, 2017.

[19] A. De Luca and L. Ferrajoli, "Exploiting robot redundancy in collision detection and reaction," in *IROS*, 2008, pp. 3299–3305.

[20] S. Haddadin, A. Albu-Schaffer, A. De Luca, and G. Hirzinger, "Collision detection and reaction: A contribution to safe physical human-robot interaction," in *IROS*, 2008, pp. 3356–3363.

[21] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast).

[22] E DIN EN ISO 10218-1:2021-09, Robots and robotic devices - Safety requirements for industrial robots - Part 1: Robot systems and integration ((prEN ISO 10218-1:2021).

[23] DIN EN ISO 12100:2011-03, Safety of machinery - General principles for design - Risk assessment and risk reduction (ISO 12100:2010).

[24] J. Vorndamme, L. Figueredo, and S. Haddadin, "Robot contact reflexes: Adaptive maneuvers in the contact reflex space," in *IROS*, 2022.

[25] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, "The role of the robot mass and velocity in physical human-robot interaction - part i: Non-constrained blunt impacts," in *ICRA*, May 2008, pp. 1331–1338.

[26] S. Haddadin, A. Albu-Schäffer, M. Frommberger, and G. Hirzinger, "The role of the robot mass and velocity in physical human-robot interaction - part II: Constrained blunt impacts," in *ICRA*, May 2008, pp. 1339–1345.

[27] N. Mansfeld, M. Hamad, M. Becker, A. G. Marin, and S. Haddadin, "Safety map: A unified representation for biomechanics impact data and robot instantaneous dynamic properties," *RA-L*, vol. 3, no. 3, pp. 1880–1887, 2018.

[28] N. Mansfeld, B. Djellab, J. R. Veuthey, F. Beck, C. Ott, and S. Haddadin, "Improving the performance of biomechanically safe velocity control for redundant robots through reflected mass minimization," in *IROS*, 2017, pp. 5390–5397.
[29] R. J. Kirschner, C. M. Micheler, Y. Zhou, S. Siegner, M. Hamad, C. Glowalla, J. Neumann, N. Rajaei, R. Burgkart, and S. Haddadin, "Towards safe robot use with edged or pointed objects: A surrogate study assembling a human hand injury protection database," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 12 680–12 687.
[30] S. Mikhel, D. Popov, and A. Klimchik, "Collision driven multi scenario approach for human collaboration with industrial robot," in *ICMRE*, 2018, p. 78–84.
[31] M. Lippi and A. Marino, "Enabling physical human-robot collaboration through contact classification and reaction," in *RO-MAN*, 2020, pp. 1196–1203.
[32] M. Lippi, G. Gillini, A. Marino, and F. Arrichiello, "A data-driven approach for contact detection, classification and reaction in physical human-robot collaboration," in *ICRA*, 2021, pp. 3597–3603.
[33] O. Sim, J. Oh, K. K. Lee, and J.-H. Oh, "Collision detection and safe reaction algorithm for non-backdrivable manipulator with single force/torque sensor," *J. Intell. Robotics Syst.*, vol. 91, no. 3–4, p. 403–412, sep 2018.
[34] E. Magrini and A. De Luca, "Human-robot coexistence and contact handling with redundant robots," in *IROS*, 2017, pp. 4611–4617.
[35] S. Yigit, C. Burghart, and H. Woern, "Applying reflexes to enhance safe human-robot- co-operation with a humanlike robot arm," in *Computer Science*, 2003.
[36] J. Kim, "Collision detection and reaction for a collaborative robot with sensorless admittance control," *Mechatronics*, vol. 84, p. 102811, 2022.
[37] A. De Luca and F. Flacco, "Integrated control for phri: Collision avoidance, detection, reaction and collaboration," in *BioRob*, 2012, pp. 288–295.
[38] E. Magrini, F. Flacco, and A. De Luca, "Control of generalized contact motion and force in physical human-robot interaction," in *ICRA*, 2015, pp. 2298–2304.
[39] C.-N. Cho, J.-H. Kim, S. D. Lee, and J.-B. Song, "Collision detection and reaction on 7 dof service robot arm using residual observer," *JMST*, vol. 26, pp. 1197–1203, 2012.
[40] Y. Kishi, Y. Yamada, and K. Yokoyama, "The role of joint stiffness enhancing collision reaction performance of collaborative robot manipulators," in *IROS*, 2012, pp. 376–381.
[41] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, "Safety analysis for a human-friendly manipulator," *International Journal of Social Robotics*, vol. 2, no. 3, pp. 235–252, 2010.
[42] R. Behrens, "Biomechanische grenzwerte für die sichere mensch-roboter-kollaboration," Ph.D. dissertation, TU Ilmenau, 2018.
[43] M. Hamad, A. Kurdas, S. Abdolshah, and S. Haddadin, "A robotics perspective on experimental injury biomechanics of human body upper extremities," in *ISR*, 2021, pp. 316–320.
[44] S. Haddadin, M. Suppa, S. Fuchs, T. Bodenmüller, A. O. Albu-Schäffer, and G. Hirzinger, "Towards the robotic co-worker," in *ISRR*, 2009.
[45] S. Haddadin, "Towards safe robots: approaching asimov's 1st law," Ph.D. dissertation, RWTH Aachen, 2011.
[46] R. I. C. Muchacho, R. Laha, L. F. Figueredo, and S. Haddadin, "A solution to slosh-free robot trajectory optimization," in *IROS*, 2022, pp. 223–230.
[47] "Pilz Robot Measurement System data sheet," https://www.pilz.com/download/open/PRMS_Set_Supplement_22225-2EN-04.pdf, accessed: 2022-03-01.
[48] ISO 9283:1998-04, Manipulating industrial robots – Performance criteria and related test methods (ISO 9283:1998).
[49] R. J. Kirschner, A. Kurdas, K. Karacan, P. Junge, S. A. Baradaran Birjandi, N. Mansfeld, S. Abdolshah, and S. Haddadin, "Towards a reference framework for tactile robot performance and safety benchmarking," in *IROS*, 2021, pp. 4290–4297.
[50] S. Haddadin, S. Parusel, L. Johannsmeier, S. Golz, S. Gabl, F. Walch, M. Sabaghian, C. Jähne, L. Hausperger, and S. Haddadin, "The franka emika robot: A reference platform for robotics research and education," *RAM*, vol. 29, no. 2, pp. 46–64, 2022.
[51] R. Laha, J. Vorndamme, L. F. Figueredo, Z. Qu, A. Swikir, C. Jähne, and S. Haddadin, "Coordinated motion generation and object placement: A reactive planning and landing approach," in *IROS*. IEEE, 2021, pp. 9401–9407.
[52] S. Golz, C. Osendorfer, and S. Haddadin, "Using tactile sensation for learning contact knowledge: Discriminate collision from physical interaction," in *ICRA*, 2015, pp. 3788–3794.
[53] DIN EN IEC 62061:2023-02, Safety of machinery - Functional safety of safety-related control systems (IEC 62061:2021).
[54] T. Tomić and S. Haddadin, "A unified framework for external wrench estimation, interaction control and collision reflexes for flying robots," in *IROS*, 2014, pp. 4197–4204.
[55] T. Tomić, C. Ott, and S. Haddadin, "External wrench estimation, collision detection, and reflex reaction for flying robots," *T-RO*, vol. 33, no. 6, pp. 1467–1482, 2017.
[56] G. Du, S. Long, F. Li, and X. Huang, "Active collision avoidance for human-robot interaction with ukf, expert system, and artificial potential field method," *Frontiers in Robotics and AI*, vol. 5, 2018.
[57] S. Haddadin, A. Albu-Schäffer, A. D. Luca, and G. Hirzinger, "Evaluation of collision detection and reaction for a human-friendly robot on biological tissues," in *IARP International Workshop on Technical challenges and for dependable robots in Human environments*, 2008.
[58] R. J. Kirschner, N. Mansfeld, S. Abdolshah, and S. Haddadin, "ISO/TS 15066: How different interpretations affect risk assessment," *arXiv preprint arXiv:2203.02706*, 2022.
[59] Universal Robots A/S, "UR5e Product Fact Sheet," https://www.universal-robots.com/media/1807465/ur5e-rgb-fact-sheet-landscape-a4.pdf, accessed: 2024-02-20.
[60] R. Bischoff, J. Kurth, G. Schreiber, R. Koeppe, A. Albu-Schaeffer, A. Beyer, O. Eiberger, S. Haddadin, A. Stemmer, G. Grunwald, and G. Hirzinger, "The kuka-dlr lightweight robot arm - a new reference platform for robotics research and manufacturing," in *ISR*, 2010, pp. 1–8.
[61] B. Proper, A. Kurdas, S. Abdolshah, S. Haddadin, and A. Saccon, "Aim-aware collision monitoring: Discriminating between expected and unexpected post-impact behaviors," *RA-L*, vol. 8, no. 8, pp. 4609–4616, 2023.
[62] (accessed 30th of August 2024) TÜV Rheinland LGA Products GmbH. [Online]. Available: https://www.tuv.com/world/en/
[63] (accessed 30th of August 2024) DEKRA Testing & Certification GmbH. [Online]. Available: https://www.dekra.de/de/testing-and-certification/
[64] (accessed 30th of August 2024) Pilz GmbH & Co. KG Inspektionsstelle. [Online]. Available: https://www.pilz.com/en-INT/services/workplace-safety/inspection-of-safeguarding-devices
[65] (accessed 30th of August 2024) SICK Vertriebs-GmbH LifeTime Services. [Online]. Available: https://www.sick.com/at/en/sick-lifetime-services/w/blog-lts-service-allgemein/

Jonathan Vorndamme received B.Sc. degrees in mechatronics in 2011 and mathematics in 2014. He received his M.Sc. degree in mechatronics in 2012. All degrees were awarded by the Gottfried Wilhelm Leibniz Universität Hannover. He started working towards a Ph.D. degree in Electrical Engineering at the Gottfried Wilhelm Leibniz Universität Hannover in 2012 and has been continuing this endeavor at the Technical University of Munich (TUM, Electrical Engineering and Computer Science) since 2018, where he is affiliated with the Munich Institute of Robotics and Machine Intelligence. His current research interests are multi-contact detection and estimation in complex robotic systems as well as robotic reflexes.

Alessandro Melone earned his bachelor's and master's degrees in Automation Engineering in 2022 from the University of Naples Federico II. His master thesis was developed during an internship at the German Aerospace Center (DLR). The thesis extends Sequential Convex Programming for dynamical systems evolving on Lie Group manifolds. He is currently a research associate at the Munich Institute of Robotics and Machine Intelligence of TUM.

Robin Kirschner studied Sports Engineering (B.Sc.) and Mechanical Engineering (M.Sc.) at Chemnitz University of Technology (CUT) with research visit at Nagoya University at the Assistive Robotics Research Group. She currently works at the Munich Institute of Robotics and Machine Intelligence (MIRMI) at TU Munich and focuses on developing concepts for physical and cognitive safety in HRI.

Luis F.C. Figueredo received his Bachelor's and Master's degrees in Electrical Engineering from the University of Brasilia, Brazil and earned his Ph.D. degree in Robotics from the same institution with an awarded PhD thesis for the biennial 2016-17. He also worked at CSAIL - MIT where he received multiple awards for robot demonstrations at venues such as IROS and ICAPS. He received the prestigious Marie Skłodowska-Curie Individual Fellowship, in 2018, for his work on biomechanics-aware human-robot interaction with AI tools acknowledged on the EU Innovation Radar, and more recently, he was also recognized within the IEEE ICRA New Generation Star Project, sponsored by NOKOV. He is currently an Assistant Professor at the University of Nottingham, UK, and an Associated Fellow at the MIRMI at TU Munich.

Sami Haddadin (IEEE Fellow) received the Dipl.-Ing. degree in Electrical engineering in 2005, the M.Sc. degree in Computer Science in 2009 from the Technical University of Munich (TUM), Munich, Germany, the Honours degree in Technology Management in 2007 from Ludwig Maximilian University, Munich, Germany, and TUM, and the Ph.D. degree from RWTH Aachen University, Aachen, Germany, in 2011. He is currently a full professor and chair of Robotics and Systems Intelligence at the Technical University of Munich (TUM), the founding director of the Munich Institute of Robotics and Machine Intelligence (MIRMI) and Vice President of Research at Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE. He has received numerous awards for his scientific work, including several best paper awards at ICRA/IROS/TRO/RA-L, the George Giralt Ph.D. Award (2012), the IEEE/RAS Early Career Award, the German President's Award for Innovation in Science and Technology (2017), and the Leibniz Prize (2019). He is a member of the National Academy of Science (Leopoldina) and the National Academy of Engineering (acatech) in Germany. He was plenary speaker at ICRA2024 and AAAI2023. He is the founder of Franka Emika (Franka Robotics under agile) and main inventor of the technology.